

THE RATIONALITY OF SCIENTIFIC DISCOVERY

PART II: AN AIM ORIENTED THEORY OF SCIENTIFIC DISCOVERY

NICHOLAS MAXWELL*

University of London

In **Part I** (*Philosophy of Science*, Vol. 41 No. 2, June, 1974) it was argued that in order to rebut Humean sceptical arguments, and thus show that it is *possible* for pure science to be rational, we need to reject standard empiricism and adopt in its stead aim oriented empiricism. **Part II** seeks to articulate in more detail a theory of rational scientific discovery within the general framework of aim oriented empiricism. It is argued that this theory (a) exhibits pure science as a rational enterprise (b) enables us to resolve problems associated with the key notions of simplicity and intelligibility (c) has important implications both for philosophy of science and for scientific practice itself.

9. Introduction. So far our almost exclusive concern has been to resolve the traditional rationality problem: How is it *possible*, in view of Humean sceptical arguments, that pure scientific inquiry can be a rational enterprise? I have argued that contemporary views of scientific inquiry are inadequate in that they all fail to solve this rather elementary problem concerning the mere *possibility* of science being rational. Aim oriented empiricism does however succeed in showing how it is possible—despite Hume—for pure science to be rational.

My concern now is with a rather more important and fruitful rationality problem, namely: How in detail ought we to pursue scientific inquiry in the best, the most rational way that we can, giving ourselves the greatest rational hope of meeting with success, granted that we accept the basic tenets of aim oriented empiricism? What, in other words, are the detailed implications for rational scientific practice, that flow from the rejection of standard empiricism, and the acceptance, in its stead, of aim oriented empiricism?

I shall argue that once we accept aim oriented empiricism we are quickly led to accepting an entirely new aim oriented theory of scientific inquiry and scientific method which has a number of fruitful implications for rational scientific practice. According to the aim oriented theory, the heart of scientific method is concerned with rationally appraising and developing different possible aims or blueprints for science. Scientific method is in essence a method of rational scientific discovery, a procedure for choosing the best possible line of inquiry in the changing circumstances of scientific research.

I shall develop this theory in a quasi-deductive fashion, by taking the basic tenets of aim oriented empiricism for granted, and then considering how we ought to pursue the specified aims of science if we are to do so in the most *rational* way possible, giving ourselves the maximum chance of meeting with success. My primary concern, then, is to spell out a *rational ideal* for scientific inquiry. The theory put forward here is not intended to mirror accurately present day scientific

* Received April, 1973.

practice. Indeed a major part of my argument is that present day scientific practice is to a considerable extent *irrational*, due to the fact that the *ideal* of scientific rationality and scientific propriety accepted by most scientists, or at least implicit in much of the institutionalized standards of scientific research, is the entirely *false* ideal of *standard empiricism*. As I pointed out in **Part I** of this essay, according to standard empiricism there cannot conceivably be a rational method of discovery, a rational procedure for choosing between rival possible aims for research, rival embryonic theories, rival blueprints. In this context of choosing between different possible lines for future research all must be hunch, guesswork, luck. According to standard empiricism, then, any attempt to choose in an articulate public fashion between rival aims for research must at best be an idle pastime, and at worst must be an utterly disreputable descent into unscientific, *a prioristic* metaphysical speculation. Thus, as a result of upholding standard empiricism and the ideal of scientific propriety that it represents, most scientists today make no attempt to choose between different ideas for future scientific research, different metaphysical blueprints, in an articulate, public, rational fashion. *Just that which aim oriented empiricism asserts to be the most important thing to do if we are to pursue science in a rational fashion becomes something altogether disreputable and unscientific once we accept standard empiricism.* In short the acceptance of standard empiricism amongst scientists leads to a positive discouragement of aim articulation and rational aim appraisal undertaken as an integral part of scientific inquiry, and this tends to inhibit successful rational scientific research. The more scientists strive to make their research conform to the false ideal of scientific propriety embodied in standard empiricism, the more sterile and unsuccessful their work must become; success in science has been achieved by those scientists who gave only a hypocritical allegiance to standard empiricism.

All this indicates that it is of some importance to spell out the detailed implications for scientific practice implicit in aim oriented empiricism. The hope is that this will enable us not only to understand science better, but will actually help us to *do* science in a better, more successful fashion. As a result of rejecting standard empiricism and accepting aim oriented empiricism in its stead, our ideal of scientific rationality and propriety will no longer inhibit and discourage scientific success, but will positively encourage us to pursue scientific inquiry in an ever more successful and rational fashion.

In presenting this new aim oriented theory of scientific method I shall in the main have *physics* in mind. The general conception of scientific method to be advocated here is however intended to apply to *all* the various sciences, and not just to physics. We shall see that on one level, the theory specifies one unified methodology for all the different sciences and, on another level, implies that different sciences—in that they have different aims—have *different* methodologies. However, in order to develop the theory in detail it will be convenient to have one particular science in mind as we proceed.

During the course of spelling out this aim oriented theory of scientific inquiry I shall come to grips with one crucial outstanding problem—namely the problem of simplicity or intelligibility.

10. The Importance of Rational Blueprint Choice for Scientific Inquiry. Our starting point, then, is that of Kepler and Einstein. At the outset we simply presuppose, without any grounds whatsoever, that the universe is ultimately simple, coherent, intelligible, harmonious, beautiful. (Or, at the very least, we assume that the world behaves *as if* it were these things, to a sufficient approximation to allow us to develop theories of ever increasing simplicity, coherence, explanatory power on the one hand, and ever increasing empirical success on the other hand.) We seek coherence, harmony, beauty, not because we have good reasons to suppose these things really do exist in the world, but because our passion to discover these things is so great, because the intrinsic *value* we place on the discovery of these things is so high, that we are prepared to devote ourselves to our long and arduous labors merely on the off-chance that what we hope to find does really exist. We seek harmony, beauty, intelligibility not as a means to some other end (such as empirical growth) but as something of *intrinsic* worth. Ugly, incoherent, unintelligible truth is of no interest to us whatsoever. If the universe is ultimately incoherent and unintelligible, then that which we seek, in pursuing scientific inquiry just does not exist and our science must founder. It is, in other words, a wild, unjustified *conjecture* that that which we seek in pursuing scientific inquiry does really exist. Our science is a shot in the dark, a wild, but rational gamble.

At first sight it looks as if science pursued in this kind of way, without any rational assurance whatsoever that that which we seek does really exist, must be grossly irrational, being no more “scientific” than the most bizarre and absurd of pseudoscientific, quasi-religious disciplines one cares to think of, such as parapsychology, scientology, or whatever. It turns out, as we have seen, that just the reverse is the case. All attempts to wean science of the massive metaphysical presupposition that the universe is ultimately simple, inevitably plunge science into irrationality. All those orthodox views which characterize science as making no metaphysical presuppositions fail to exhibit science as a rational enterprise in that these views fail to provide any rationale for the exclusion of empirically successful *aberrant* theories.

In pursuing scientific inquiry our ultimate aim, then, is to convert a more or less vague metaphysical theory—which asserts that the universe is ultimately simple, coherent etc.—into a precise, fully articulated, empirically testable scientific theory. More modestly, our aim is to develop theories which are progressively more and more successful empirically and which, at the same time, succeed in articulating more and more of the presupposed metaphysical theory. Our presupposed metaphysical theory is a kind of blueprint for our ideal future scientific theory.

I come now to a point which I cannot emphasize too strongly. In setting up our science, and in pursuing the study of our science, *the single most important and most intractable problem that can face us is precisely the problem of discovering the best possible aim or metaphysical blueprint for our science.*

It is hardly too much to say that *everything* depends on making the appropriate choice of aim or metaphysical blueprint. If we choose well, and our blueprint does correspond *at least roughly* to how things really are, then there is every chance that our science will forge ahead with great empirical success. For in this case our blue-

print will guide us to consider only appropriate kinds of new theories. And if in addition to being more or less true, our blueprint is relatively *precise* and *specific*, then we will have available all the more restrictive a delimitation of appropriate new theories to be considered; we will have something approaching a *method of discovery*.

But if our fundamental metaphysical blueprint is a grotesquely *false* theory, so that the world—although intelligible—is not intelligible in anything like the kind of way specified by our blueprint, then our science *cannot succeed*. However rationally and brilliantly we pursue our science, we can get nowhere since all the time in considering theories that are compatible with our basic blueprint, we will be considering the wrong *kind* of theory, we will be considering theories that are all inevitably completely false.

It is thus absolutely crucial for the success of our science that we make the best possible choice of blueprint. The history of science is in fact littered with research programs which were pursued with immense creative energy and “rationality,” but which, after perhaps some initial success, got nowhere because the wrong choice of blueprint had been made, and this wrong choice was held onto inflexibly. One need only think of the research program of Plato and Ptolemy designed to show that the motion of heavenly bodies can be reduced to uniform circular motion (a blueprint retained even by Copernicus!). Or alternatively, one might think of the Aristotelian program for physics. One absolutely decisive factor, then, in the success achieved by such men as Kepler, Galileo and Newton, was—not that they repudiated metaphysics—but that they had the luck to choose a metaphysical view, a blueprint, which is near enough to how things really are—or at least near enough to how things appear to be. The thesis that the book of nature is written in the language of mathematics, that simple precise mathematical laws govern the motion of bodies, is a vast, wild conjecture, which today we all take for granted because of the immense empirical success of physics based on this conjecture. There is however no *a priori* reason why the universe has to be like this. The universe might very well be intelligible—in terms perhaps of some overall cosmic purpose—even though phenomena cannot be precisely described in terms of simple mathematical laws. And if this had been the case, the investigations of Kepler, Galileo and Newton would have got nowhere!

The most important problem of all, then, that faces us in pursuing our search for intelligibility is to make the best possible choice of metaphysical blueprint. *But it is just here that our science is most speculative, most vulnerable, most open to error!* We have no reason whatsoever to suppose that the universe *is* intelligible at all. Much less, then, do we have a reason for supposing that the universe is intelligible in precisely such and such a way (e.g. because everything is made up of infinitely hard corpuscles which interact only by contact, or because everything is designed to fulfill some overall cosmic master plan). We can have, it seems, no reason whatsoever for supposing that some specific metaphysical blueprint is actually true. Our chances of hitting on anything like the true blueprint seem infinitely remote.

Thus that issue which it is of the utmost importance for us to settle correctly for our whole science—namely the issue of choosing a more or less true blueprint—is

also the issue which is the most difficult to settle correctly. Our choice of blueprint must remain irredeemably speculative and conjectural, an all but blind shot in the dark.

It might be thought that these considerations simply show that no science can be rational which does make metaphysical presuppositions, and hence that the lesson to be learnt from all this is that we should forgo making all such unfounded metaphysical presuppositions. But this we cannot do. As we saw at length in **Part I**, it is precisely the overt disavowal of metaphysical presuppositions which creates the traditional rationality problem. Our fundamental aim is to seek intelligibility, comprehensibility; we must, then, if we are to be rational, at least *conjecture* that what we are looking for does exist; and hence we cannot avoid the problem of trying to decide what particular *kind* of intelligibility exists in the world.

There is another consideration which reinforces this point. *Our whole methodology* (i.e. our set of rules of theory-acceptance) *depends on our choice of metaphysical blueprint*. If, for example, we choose as our basic blueprint some such thesis as "Everything that occurs does so in order to fulfil some overall cosmic purpose," then the whole methodology of our science will be radically different from what it would be if we chose a blueprint such as: "Everything that happens is a consequence of the interactions of fundamental corpuscles." In the two cases, entirely different kinds of theories would be considered acceptable. In other words, our methodological rules of theory-acceptance, our criteria of intelligibility for theories, would be quite different. Thus our whole methodology depends on our choice of blueprint, and if we are to choose our methodology rationally it is obligatory that we choose *rationally* our fundamental metaphysical blueprint. And if we try to forgo choosing any blueprint at all, then we are bereft of any methodology, we can have no criteria of intelligibility, and we could not proceed at all. (The infinite realm of *aberrant* theories could not be rationally excluded.)

All this shows that we cannot adopt the standard attitude to the problem of choosing the best metaphysical blueprint. We *cannot* argue that since choosing any particular metaphysical theory as the aim of science must be so utterly speculative and open to error, that it would be better not to choose any particular metaphysical blueprint at all. We have no alternative but to make some kind of choice.

Thus the conclusion to be drawn from the fact that it is extremely difficult to make the best possible choice of blueprint is just the opposite of the orthodox viewpoint. Because choosing the best blueprint is both *supremely difficult and supremely important*, it is just here, above all, that we need to be as rational, conscious, deliberate and careful as possible. Instead of trying to brush the whole issue under the carpet, by leaving the matter to the hunches of individual investigators, we need to make the problem of choice of blueprint, choice of aim, as explicit, as open, public and objective as we possibly can. It is only if we choose our blueprint *rationally* that we can, it would seem, stand the faintest chance of succeeding with our science.

We reach, then, a very remarkable result. In **Part I** it emerged that the fundamental *epistemological* or *methodological* problem is the problem of the *aims* of science. The major hurdle in the way of resolving the pure rationality problem was

just explicitly acknowledging that science seeks explanatory truth, intelligible truth and thus makes massive metaphysical presuppositions. It now turns out that this basic *epistemological* problem is *also* the basic *scientific* problem. The fundamental problem that needs to be resolved, if we are to pursue scientific inquiry in the best, the most rational way possible is just: *How are we to make the best, the most rational choice of aim, of metaphysical blueprint?* The aim oriented theory of scientific inquiry to be given here is designed essentially to answer this basic scientific rationality question. The methodological rules of science will turn out to be rules for the rational appraisal of possible aims, or metaphysical blueprints construed as aims of science.

Before we consider in detail the central problem of how the best possible aim for a science can be chosen, there is one simple but absolutely crucial point that must be stressed first. The point is this. If we are to choose rationally the best possible aim for our science it is essential that we consider alternative possible aims, it is essential that we remain constantly prepared to modify, or even drastically change the aim that we have chosen. The chances that, at any stage in our inquiries, we have made the best possible choice of aim are slim indeed. If we are to proceed rationally, then, it is essential that we attempt constantly to improve the chosen aim, or metaphysical blueprint, of our science.

It might be thought that this readiness to change and improve the aim of a science is important only during the early stages of a science, before much empirical success has been achieved, but that once a science has become as mature and successful as modern physics, let us say, there is no longer any need to consider whether the aims of the science should be improved, and thus changed.

This conclusion is quite wrong. However successful a science may be, the problem of *improving* the aim of that science will always remain perhaps the most fundamental and important problem of that science.

This point can be brought out like this. In choosing the best possible aim, or metaphysical blueprint, for a science, one invariably has to make the following kind of agonizing compromise between two opposing sets of desiderata.

On the one hand, it is highly desirable to choose an aim or metaphysical blueprint which is as "broad," as loosely defined, as unrestrictive as possible. For if one chooses a very precise, narrowly-defined blueprint, one makes it very likely indeed that the blueprint is *false*, and that the whole science is committed to a blind alley. By choosing an aim as broad as possible, one does not commit one's science to too many highly specific metaphysical presuppositions, and hence there is much more chance that one's science has not been committed to a blind alley. Another way of putting this point would be to say that one should always have available a wide spectrum of different possible precise blueprints, instead of choosing just *one* of these precise blueprints, thereby risking that an entirely *false* blueprint has been chosen.

According to this kind of consideration Plato should have chosen for astronomy the relatively broad blueprint: "Heavenly bodies move in accordance with precise simple mathematical laws" (which leaves open a wide choice of more specific blueprints). In choosing the highly specific blueprint: "Heavenly bodies move in circles

with uniform motion” Plato chose a much too *specific* blueprint, and thus committed astronomy to a blind alley. Again, according to this kind of consideration, in the seventeenth century it would be preferable to choose Galileo’s blueprint: “Phenomena occur in accordance with a few simple, precise mathematical laws”; rather than the much more precise rendering of this blueprint: “The world is made up of a few different kinds of infinitely hard corpuscles, which interact only by contact, and which move in accordance with the laws of mechanics.”

All this, however, only spells out one kind of desideratum to be borne in mind in choosing the best possible aim available. For, on the other hand, a rather different set of considerations suggest that it is preferable to choose the most *precise*, the most *restrictive*, the most narrowly defined blueprint possible. If we choose a very broad, loosely defined aim, we will have little help available when it comes to developing specific testable theories put forward in an attempt to realize the aim. The more imprecise a blueprint is, the wider will be the range of possible testable theories compatible with the blueprint; and as a result the process of putting forward new conjectures—candidates for serious scientific consideration—will be all the more random, arbitrary and difficult. If, however, a science has a very *precise*, narrowly defined blueprint, then possible candidates for future scientific theories will be much more restrictively defined. Only a highly specific *type* of theory will be suitable. And if the blueprint is very precise and specific, there is always the hope that by making it just a little more precise—perhaps in an entirely nonarbitrary way—the blueprint will itself be transformed into a fully fledged testable scientific theory. In this way the process of articulating and improving the aim of the science will of itself have led to the development, or to the discovery, of a new scientific theory!

Thus, from the standpoint of developing new scientific hypotheses, we need a blueprint as precise and narrowly defined as possible. We need one highly specific blueprint, and not a whole *range* of such blueprints. From this standpoint, Plato was entirely rational to advocate that astronomy adopt the highly specific blueprint: “Heavenly bodies move at a uniform rate in circles.” Seventeenth century philosophers were entirely rational to choose some highly specific version of Galileo’s blueprint, e.g. some version of the corpuscular hypothesis.

It is clear that if we are to choose the best possible blueprint available, if we are to make our choice in the most *rational* way possible, then we must somehow make a judicious compromise between the above two opposing kinds of desiderata. We must pick a blueprint that is as specific, as narrowly defined as possible, so that the field of possible new scientific theories are narrowed down as much as can be, and we have something approaching a rational method of discovery. But on the other hand we must be constantly alive to the possibility that we may have committed our science to a false blueprint, to a blind alley, and we must be ready, particularly when our theorizing meets with no empirical success whatsoever, to *broaden* our blueprint, to modify it, so that a wider range of possible new theories can be considered.

It is, I hope, clear that, from a rational standpoint, one would expect the need for this constant, agonized reappraisal of one’s basic aim or blueprint to persist as a science evolves, however empirically successful and “mature” that science may

become. For if a science is highly successful empirically, if successive theories are put forward which lead to great empirical growth, and which move towards giving a more and more complete articulation of some rather precise blueprint, then one will want to make this “best” blueprint even more precise, in the hope that one will thereby transform it into a testable scientific hypothesis. Thus, however empirically successful our science may be, it will always be desirable, from a rational point of view, to narrow down the blueprint as much as possible, thereby strengthening the “logic of discovery” aspect of our science; and hence we will always need to be alive to the possibility that the blueprint has been made so specific that it is in fact *false*, and that further progress requires a *broadening* of the blueprint. We will need constantly to reassess the blueprint that we have chosen, in an attempt to pick the best possible blueprint available for our science.

There is an extremely important *additional* reason for claiming that, if we are to proceed in a rational fashion, then we will need constantly to modify, to *improve* our metaphysical blueprint. This additional reason can be put like this.

Our prime objective, in pursuing our science, is to seek *explanatory* truth, *intelligible* truth. Our science places a premium on intelligibility, simplicity, beauty *for its own sake*, and not as a means to some other end. We are therefore rationally entitled, other things being equal, to choose that metaphysical blueprint, *M*, say, which embodies our very highest ideal of intelligibility, simplicity, beauty, which promises to lead us to the development of the best possible explanation of phenomena.

However, having made such a choice, it is more than likely that as we proceed we will begin to notice certain purely *a priori* imperfections in *M*. *M*, we will realize, contains certain wholly arbitrary assumptions. *M* may be simply a special case of a much more general kind of metaphysical theory. It may even be the case that we discover inconsistencies in *M*. In brief, as we proceed, we discover that *M* can no longer be taken as embodying our ideal of intelligibility, simplicity. Our standards of intelligibility have been raised. Our aim, our blueprint, for our science, needs to be *improved*. We need a new metaphysical theory *M*¹ to replace *M*. And granted that we can establish, in an entirely objective fashion, that *M*¹ embodies greater simplicity, intelligibility than *M*, then, other things being equal, it will be entirely rational for us to adopt *M*¹ as the aim of our science rather than *M*, since our whole science places an intrinsic premium on intelligibility, for its own sake.

Thus it is to be expected that as we proceed we will need constantly to *improve* our basic metaphysical blueprint as we detect inadequacies in the existing blueprint from the point of view of intelligibility, or simplicity. Our standards of intelligibility or simplicity, our conception of what can be “explained,” ought constantly to rise, and become more ambitious, if we are to seek for the best possible explanation for phenomena in the most rational way. It is my view and I shall argue that through much of the history of *physics* we can see a constant *raising* of the standards of intelligibility, a constant improvement in the fundamental blueprint of physics from the standpoint of intelligibility considerations.

The model of scientific inquiry articulated so far, might be schematized like this (see Figure 1).

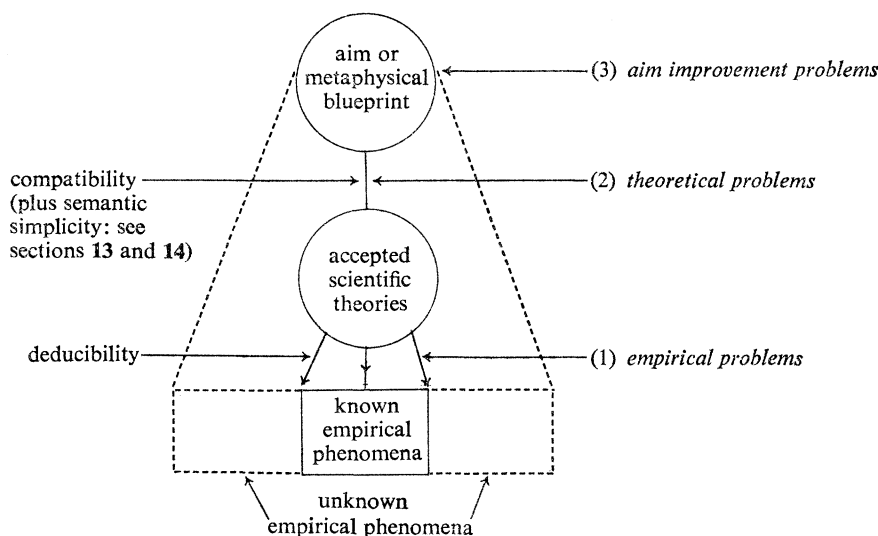


Figure 1

The aim of science is to develop testable scientific theories which give a more and more precise and complete articulation of the metaphysical blueprint, and which at the same time successfully predict more and more empirical phenomena. The thesis that the universe is ultimately simple, coherent, intelligible (in some more or less specific way) is just the thesis that a physical theory which gave a complete, precise articulation of the presupposed metaphysical blueprint, would in principle predict all physically possible empirical phenomena (in the diagram I have tried to represent this by means of the slanting dotted lines). At any stage at least three kinds of problems should arise in a science. First, there will be *empirical problems*, which arise when the accepted scientific theories clash with (or at least fail to predict) accepted empirical phenomena. (In this case either the relevant theories need to be modified or replaced, or the empirical “laws” need to be modified or replaced. We cannot assume that in the advent of a clash between theory and experiment it is always the theory which must give way.)

Second, there will be what I have called *theoretical problems* which arise essentially because accepted scientific theories do not succeed in providing anything like a complete, precise articulation of the presupposed metaphysical blueprint. Theoretical problems will arise even in the absence of serious empirical problems. A theoretical problem, typically, will take the following form: How can the current, best, fundamental theory, *T*, be so modified or replaced that a more complete articulation of the metaphysical blueprint may be achieved in the form of an empirically testable scientific theory? One specific kind of theoretical problem arises when two fundamental scientific theories clash (as in the case today, for example, of the general theory of relativity and quantum mechanics). An absolute minimal demand for an adequate scientific articulation of the presupposed metaphysical blueprint is, after all, that the accepted scientific theories are *compatible*.

Another specific kind of theoretical problem arises when an empirically successful fundamental theory is more or less *incompatible* with the adopted aim, or metaphysical blueprint, of the science. In this case one of two things may happen: either the theory is retained and the aim, the metaphysical blueprint is modified, so that it becomes compatible with the theory; or alternatively, the metaphysical blueprint is retained and the *theory* is modified or replaced. We will be rational in attempting to bring the latter about (i.e. in attempting to make our theory conform to our aim, rather than our aim to our theory) in those situations where to change our aim would lead to a drastic decrease in the inherent simplicity, intelligibility etc. of our adopted metaphysical blueprint. Thus, just as accepted empirical "laws" may quite rationally be rejected, or revised, when they clash with a "good" theory, so a good theory may quite rationally be rejected or modified if it clashes with a good aim.

The third kind of problem that arises in scientific inquiry, according to the model being advocated here, arises in connection with attempts to improve the accepted aim, or metaphysical blueprint. As I have already remarked, a metaphysical blueprint may initially appear to embody the quintessence of simplicity, coherence and intelligibility; but may later be found to be in certain aspects arbitrary, incoherent, or even inconsistent. Aim improvement problems arise when such arbitrary, incoherent or inconsistent features are discovered in the accepted metaphysical blueprint.

Once we accept aim oriented empiricism and the above aim oriented model of scientific inquiry, the very serious and damaging inadequacies of orthodox views about science become all too apparent. In contrast to the above aim oriented model of scientific inquiry schematized in Figure 1, the more sophisticated versions of the orthodox model might be schematized like this:

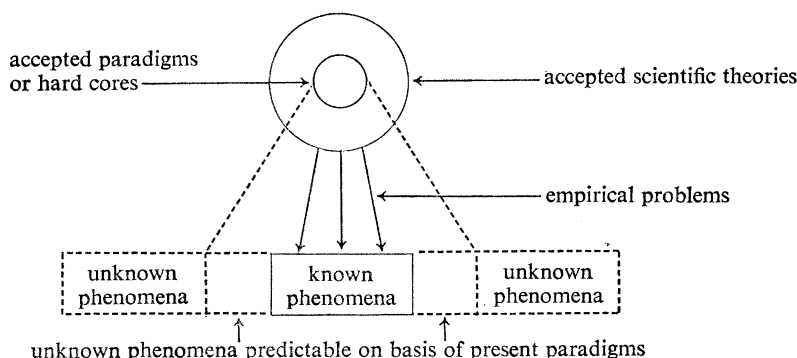


Figure 2

The difference between Figures 1 and 2 is quite startling. The problems that are perhaps the most important from the point of view of the development of knowledge—namely what I have called *theoretical* problems and *aim improvement* problems—have simply vanished from the picture. The orthodox model, in suppressing the metaphysical presuppositions of science, suppresses all hope of

rationally assessing and rationally improving these presuppositions. And perhaps *most* disastrously of all, the orthodox model suppresses any possibility of developing new theories, in a rational manner, by articulating and making more precise the metaphysical presuppositions.

11. An Aim Oriented Theory of Scientific Method. I come now to the specific theory of scientific method that I wish to advocate.

The fundamental problem of scientific inquiry, at any stage, is, I have argued, to make the best possible choice of aim, of metaphysical blueprint. This is both the most *important* thing to get right, and often the most *difficult*. The fundamental *rules* of scientific investigation are, then, I suggest, rules for selecting the best possible aim or blueprint in various sorts of circumstances. A mature science ought, I suggest, to proceed by steadily and progressively modifying and improving the chosen blueprint in a highly conscious, articulate, public, objective manner, by solving aim articulation problems inherent in the chosen blueprint, and by modifying the chosen blueprint in response to empirical pressure, in a highly *critical* fashion. As we shall see below, a succession of rational modifications in a given blueprint, designed to solve successive problems of the blueprint, can lead to dramatic, revolutionary changes at the level of scientific theory. In other words, a steady, rational continuity of development is often discernible in science at the level of blueprints, where all is discontinuity and revolution at the level of theory. Thus, whereas methodological rules which specified how one ought to modify and improve scientific *theories* in progressive almost continuous steps would quickly stultify scientific development (since vital discontinuous revolutionary steps would be repressed) this does not apply in the case of *blueprints*. Steady, rational blueprint modification can lead to profound and dramatic conceptual change. The methodological rules put forward here are then quite properly primarily rules for the *modification* of existing blueprints.

Our discussion of the problem of how to make a rational choice of the best available blueprint, in section 9, has in effect already provided us with a few elementary rules for the selection of the best aim. We have to begin with:

Rule 1. Other things being equal, choose that aim, that blueprint, which is the most intelligible, simple, coherent, harmonious, explanatory, unified, beautiful. (In part at least this will mean choose that blueprint which promises to lead to the development of the most intelligible, simple etc. testable scientific theory.)

The rationale behind this rule is quite simply that our whole investigation places a premium on simplicity, intelligibility, *for its own sake*. We thus quite reasonably choose that blueprint which gives the best promise of leading us to the realization of our overall aim.

We also saw in the last section that in picking the best possible aim we must make a kind of agonized compromise—a compromise which will need constant reappraisal—between picking the narrowest possible blueprint—and hence having something approaching a logic of discovery—and picking the broadest possible blueprint—in which case we lessen the danger of committing our science to a blind alley. We need, then, two rules which specify in what circumstances it is rational to

make our chosen aim narrower or broader. These two rules may perhaps be put like this.

Rule 2. Suppose the aim of a science is the rather broad blueprint B (which we may think of as being the conjunction of the different, much more precise blueprints $B_1, B_2 \dots B_n$). Suppose attempts to realize B have led to the development of a succession of theories $T_1 T_2 \dots$ which have been enormously successful empirically, i.e. have led to great empirical growth. Further, let us suppose that these successive theories can very reasonably be held to give successively more complete articulations of B , say, but cannot be held to tend towards any of B_1 or $B_2 \dots B_n$ excluding B . The empirically very successful research program points at B , as it were, and does not point at any of B_1, B_2 etc. In these circumstances, other things being equal, it will be rational to choose B as the aim of the science rather than B .

The next rule specifies the circumstances under which we ought to broaden our aim.

Rule 3. Suppose the aim of a science is B , and theories put forward in attempts to realize B have met with little or no empirical success. In this case, other things being equal, it will be rational to *broaden* B , so that it becomes, let us say, the conjunction of B_1 and B .

In brief, *Rule 2* tells us that if our science is meeting with great empirical success, then we are entitled to narrow our blueprint, if we can do this in a nonarbitrary way. And *Rule 3* tells us that if our science is meeting with little or no empirical success then we are entitled to broaden our blueprint. But of course this narrowing and broadening of aims must be done, as much as possible in conformity with *Rule 1*.

There are now two further general rules of scientific inquiry which deserve to be mentioned.

Rule 4. Other things being equal, choose that aim which gives the best *a priori* promise of leading to an empirically successful research program.

Consider the following two radically different blueprints: M_1 : all phenomena occur in accordance with simple, precise mathematical laws. M_2 : all phenomena arise as a consequence of the desires, intentions, aims, of spirits which inhabit the world. The point is now this. If M_1 is true, we have every hope of developing an immensely successful and productive empirical research program, taking M_1 as the basic blueprint. There is every hope of developing theories which have successively greater and greater explanatory power, which provide an ever more comprehensive, more precise account of phenomena while at the same time moving towards providing a steadily more complete articulation of the basic blueprint. There is in other words every hope that the kind of *progressive development* that is such a striking characteristic of physics during the last three hundred years, will take place. Early theories (Galileo, Kepler) will be corrected and surpassed by later theories (Newton), which will in turn be corrected and surpassed by later theories still (Einstein), but at the same time the later theories will reduce, as an approximation within some restricted domain, to the earlier theories. Even though corrected and surpassed, the early theories will remain permanent contributions to the development of knowledge.

If, however, we suppose M_2 is true, all this looks most unlikely. Any number of particular, *ad hoc* postulates will be needed concerning the desires, moods, whims, intentions of all the various spirits, and it seems unlikely that a research program based on M_2 could develop the *progressive* character just described. Particular hypotheses concerning spirits even if true for a while, may well become out of date as spirits change their goals. Knowledge will constantly need to be revised, without this endless revision necessarily leading to an overall *progressive evolution*, or *development*.

Other things being equal, then, we ought to choose M_1 rather than M_2 , since M_1 offers more hope of success than M_2 . It was, I suggest, primarily this consideration which favored—from a rational standpoint—the metaphysics of Galileo, and of the seventeenth century corpuscularians over the metaphysics of Aristotelianism. Galileo's blueprint, if true, promises great empirical growth, whereas the blueprint of Aristotelianism with its *ad hoc* and teleological characteristics, if true, promises only rather slow, piecemeal, unprogressive empirical growth.

The example I have chosen to illustrate *Rule 4* may seem today rather fanciful. It would be wrong however to conclude from this that *Rule 4* has no relevance for contemporary science. I am inclined to think that it plays a decisive factor in the selection of aims or blueprints for scientific research projects. Invariably, a scientist will be confronted by innumerable, possible alternative blueprints, none of which may be ruled out on grounds of implausibility. If he is to proceed at all, it is vital that he has some tentative *a priori* way of grading these possible blueprints. One way is, of course, to grade them in terms of inherent intelligibility or simplicity. But another extremely important way is to grade them by means of *Rule 4*. To outsiders scientists often seem incredibly myopic and conservative in that unconventional metaphysical ideas—such as those of parapsychology—are simply ignored. But from the standpoint of science the problem is to *cut down* on the vast wealth of possible metaphysical blueprints; and given the fundamental concern of science to develop empirically progressive research programs, it is essential—and certainly entirely rational—that science should favor those metaphysical blueprints which *promise* to lead to the development of empirically progressive research programs.

We come now to a rule of a rather different type from the above. It is a rule which specifies how we ought to choose ordinary methodological rules of theory-acceptance once we have chosen the best possible aim. In essence the rule can be stated quite simply like this.

Rule 5. Choose those methodological rules of theory-acceptance which give the best, the most rational hope of leading to a successful realization of the best aim, or metaphysical blueprint.

The decisive point behind *Rule 5* is this. The kind of theory that we will regard as acceptable may well depend crucially on the kind of aim we have adopted, the kind of metaphysical blueprint which exemplifies our ideal of an intelligible, comprehensive theory. Thus if our blueprint is: "The world is made up entirely of many infinitely hard corpuscles which interact only by contact," the kind of theory that we will be prepared to regard as intelligible and acceptable will in general be quite different from what we would regard as intelligible and acceptable if our blue-

print were, let us say: “Everything occurs in order to help bring about some overall cosmic purpose.” The two blueprints would give rise to different methodological rules of theory-acceptance. If we adopt the first blueprint, then, quite rationally we would adopt methodological rules such as: Accept action-by-contact theories. Reject action-at-a-distance theories. Reject teleological theories. If, on the other hand, we were to adopt the second blueprint, then we should of course accept the methodological rule: Only teleological theories are ultimately acceptable.

Thus the methodological rules, *R*, of theory-acceptance, which it is rational to accept for a science, depend crucially on the nature of the blueprint *B* of that science. A *change* of blueprint *B* is almost bound to necessitate a change of methodological rules. Thus at once we see the need for *Rule 5*.

The thing might be summed up like this. Once we accept the above aim oriented model of scientific inquiry, we are led immediately to the thesis that a science can be rational only if it is ready constantly to *change* its basic blueprint. And this in turn immediately leads to the conclusion that the science must be ready constantly to change its methodological rules of theory-acceptance. And this in turn implies that we need a rule, such as *Rule 5*, which specifies how methodological rules of theory-acceptance ought rationally to be changed with a change of blueprint.

We need to distinguish here three levels of statements,

level 1—ordinary scientific theories.

level 2—methodological rules which specify under what circumstances scientific theories (level 1 statements) should be accepted and rejected (or how they should be graded).

level 3—methodological rules which specify under what circumstances level 2 rules should be accepted and rejected.

The above *Rules 1–5* are, in fact, all *level 3 rules*, since in specifying what *aim or blueprint* ought to be chosen, these rules in effect determine—*via Rule 5*—what level 2 methodological rules ought to be chosen. The theory of scientific method embodied in the above methodological *Rules 1–5* thus constitutes an entirely new *conception* of scientific method. For up until now all those who have sought to characterize scientific method have tried to do so in terms of level 2 methodological rules.

The reason why it is absolutely obligatory to characterize scientific method on what up until now has been thought of as the “metamethodological” level—i.e. in terms of level 3 rules—can be brought out quite decisively from the following elementary considerations. As we saw in section 5 of **Part I**, whether or not it is rational to accept or reject a given theory may depend crucially on what our *purpose*, our ultimate aim is, in considering the theory. In other words *different aims give rise to different rational rules of theory acceptance*. Now if all the different sciences at all times had one and the same basic aim—to search for truth *as such*, let us suppose—then the fixed, universal rules of scientific inquiry could well be formulated entirely in terms of level 2 statements. We would only need to move onto the “metamethodological” level to consider the question of the rationality or acceptability of any proposed level 2 rules; and once agreement had been reached,

then scientific method could be quite adequately characterized in terms of the agreed level 2 rules of theory-acceptance.

The basic assumption here—that all the different sciences at all times have the same fixed *aim*—is, however, quite grotesquely *false*. First, as we have seen, a single science, from a purely *rational* standpoint ought constantly to change its basic aim or blueprint as it progresses. And if we actually look at the development of any science we see a constant change and evolution of basic blueprint. Physics, for example, can be regarded as having had in succession the following five very different basic aims or metaphysical blueprints.

- (a) Aristotle's teleological metaphysical system, which asserts that phenomena occur as a consequence of things trying to *actualize* their *potentia*.
- (b) The aim of seventeenth century corpuscularians and Cartesians to explain phenomena in terms of the motions of corpuscles which interact only by contact.
- (c) The aim of classical mechanistic physics (Newton (?), Boscovich, Helmholtz) to explain all phenomena in terms of the motion of point-atoms which interact by means of central forces of attraction and repulsion which vary with distance.
- (d) The aim of classical field theory (Einstein): to explain all phenomena in terms of a unified field theory.
- (e) The aim of modern physics, which is highly problematic, but which differs radically from (a) to (d) in that nature is assumed to have a fundamentally *indeterministic* or *probabilistic* feature.

There is of course nothing special about physics in this connection. Almost all other sciences have experienced radical changes of basic *aims* as they have evolved.

It might be thought that if we consider the successive blueprints of physics indicated above ((a) to (e)), then we can detect certain features that are common to *all* these blueprints, and these common features give rise to *fixed* level 2 rules of theory-acceptance in physics. It is, for example, often held that acceptable physical laws must be strictly universal, must refer to no specific time or place, or must be *invariant* with respect to space and time. But the idea that there can be such fixed level 2 intelligibility rules of theory-acceptance, which cannot be abandoned within the context of physical inquiry, is in effect to commit physics to a dogmatic, irrational aim. To demand that physical laws be invariant with respect to space and time is to commit physics to a specific metaphysical blueprint—which postulates homogeneity to space and time—which cannot be rejected or revised in any circumstances whatsoever. And this is clearly undesirable. Nature may well be intelligible even though her basic laws are not invariant with respect to space and time—and if we are to proceed rationally, we certainly ought not to exclude *a priori* such a possibility.

As far as physics is concerned, the idea that there are in the world certain *invariant* properties which persist through all change no doubt lies at the very heart of the idea that Nature is intelligible. Indeed the blueprints (a) to (e) above should be interpreted just as specifying what are the *unchanging* properties of things. We

want to be able, however, to change our views as to what are the ultimate invariant properties in the world. And to do this will in effect involve changing the level 2 intelligibility rules of theory-acceptance. Modern physics in fact reveals that the accepted level 2 intelligibility rules—invariance and symmetry principles—do change with time. Even the time honored rules concerning the invariance of laws with respect to space and time—closely associated with the blueprint thesis that space and time are homogeneous—were denied by Aristotle, and have been radically transformed by general relativity. General relativity can hardly be said to conform to these rules, since it is itself a theory about the structure of space-time. Rather, general relativity may be said to conform to a generalization of these invariance principles, namely to the principle of general covariance.

My fundamental argument in support of the general theory of scientific method being advocated here can, then, be summarized like this.

- (i) Once we accept the general aim oriented viewpoint it becomes clear that the single most important and most intractable problem that faces a science is to make the best, the most rational choice of basic blueprint. The basic *rules* of scientific inquiry need then to be rules for the rational selection of the best possible blueprint in the ever changing context of scientific research, and it is just this which the theory of scientific method advocated here supplies.
- (ii) Once we accept (a) that the acceptability of a theory depends on our basic *aim*; and (b) the rational pursuit of scientific inquiry requires a constant *change* of aim or blueprint, then it becomes clear that level 2 rules of theory-acceptability will change as science evolves. It becomes inevitable that the fixed, universal rules of scientific inquiry must be formulated as *level 3 rules*, which specify under what circumstances the aim of a science should be changed, and how such a change effects accepted level 2 rules. We thus end up with the theory of scientific method advocated here, and summarized in terms of the above level 3 *Rules 1–5*.

There are now one or two brief comments that I wish to make about the above theory of scientific method.

It should be noted first that *Rules 1* and *4* are both *a priori* rules, in that they specify criteria for the selection of blueprints in terms of nonempirical considerations. *Rules 2* and *3* specify how blueprints ought to be modified or changed in the light of empirical considerations. And *Rule 5* in effect specifies what it *means* to say that an empirically progressive succession of theories T_1, T_2, \dots favors or does not favor some particular blueprint B on empirical grounds. If the theories T_1, T_2, \dots are increasingly *acceptable* from the point of view of those level 2 rules most appropriate for the blueprint B , then B is favored on empirical grounds. If, however, the theories T_1, T_2, \dots are increasingly *unacceptable* from the point of view of level 2 rules appropriate to B , then B meets with corresponding empirical disfavor.

It is more than likely that the *a priori Rules 1* and *4* will clash with the empirical *Rules 2* and *3*, in that *Rule 3*, let us say, favors a blueprint B_1 and *Rule 1* favors a rather different blueprint B_2 . Such a clash in effect gives rise to the kind of theoretic-

cal problem discussed briefly above. No foolproof, general rule can be laid down for resolving such a clash. One can say, however, that in attempting to *improve* the most intelligible blueprint by making it even more intelligible (in accordance with *Rule 1*), by solving what I have called aim improvement problems, it will be extremely important to take up hints and suggestions from the most empirically successful lines of research. If the most intelligible blueprint is B_1 , and the most empirically successful blueprint is B_2 , then every attempt should be made to see whether B_1 can be improved on intelligibility grounds so that, in addition, it becomes compatible with B_2 . In other words, in the advent of a clash between *Rules 1* and *4* (which favor B_1) and *Rules 2* and *3* (which favor B_2), every attempt should be made to evolve a new aim B_3 which is favored by all four rules. Aim improvement on intelligibility grounds may well need the constant inspiration of empirical considerations; it should not however be the *slave* of empirical considerations. It is, of course, just the fundamental metaphysical presupposition inherent in scientific inquiry that *Rules 1, 2, 3* and *4* will not permanently clash. If a science develops in such a way that the blueprint favored by *Rule 1* becomes ever more and more different from the blueprint favored by *Rules 2* and *3*, then that science becomes increasingly *problematic*. A science in such a state ceases to be genuinely progressive and successful, even though empirical success may still be achieved.

The theory of scientific method embodied in the above *Rules 1–5* is primarily a methodology which assesses the acceptability of theories. But in order to do this rationally, we need to assess possible aims or blueprints, and hence the emphasis on the assessment of blueprints in the above rules. The theory is also, however, implicitly, a methodology of scientific *discovery*, a methodology for the rational development or invention of new scientific hypotheses. In order to make this aspect of the theory explicit, here is a sixth rule, more or less implicit in *Rules 1–5*, which specifies procedures for the rational development of new scientific theories.

Rule 6. New scientific theories are to be developed by detecting and solving the three kinds of problems indicated above. Of these, the most important by far are aim improvement problems. By articulating more and more clearly the chosen blueprint of the science, and seeking ways of making the blueprint more and more intelligible, simple, precise and restrictive, in a nonarbitrary way, there is the hope that one will be able to formulate level 2 methodological principles (such as the various invariance and symmetry principles of modern physics), which any new theory must conform to if it is to help us realize the chosen blueprint. Such level 2 principles of intelligibility or simplicity may then lead to the development of new scientific hypotheses. As we shall see below, it was precisely this rational method of discovery that Einstein employed in developing the special and general theories of relativity.

Identifying and solving theoretical or compatibility problems also provides a rational method for the discovery of new theories. Here one might mention Dirac's development of his relativistic quantum theory, evolved precisely as a solution to the problem of making quantum mechanics compatible with the level 2 principles of intelligibility or simplicity of Lorentz invariance.

Finally, in order to develop new and better scientific theories, it is of course

important to detect and attempt to solve empirical problems. An empirical problem that cannot be solved within the framework of the existing scientific theory *T* alerts us to the fact that *T* is unsatisfactory and needs to be replaced. In other words, we discover that *T* is not an adequate solution to the aim improvement or compatibility problem which it was designed to solve; we therefore need to return again to these problems to develop a more adequate solution.

The discovery of empirical problems is a vital part of the overall process of scientific discovery and development. However, according to the theory advocated here, empirical problems are not the sole spur to scientific development. Both aim improvement problems and compatibility problems can exist in the complete absence of any empirical problems; and when it comes to the development of fundamental new theories, aim improvement and compatibility problems are without doubt far more important and fruitful than empirical problems. It is, in other words, by tackling aim improvement and compatibility problems that fundamental advances are made in a subject; empirical problems only become important in this respect if there are reasons for holding that such problems are closely associated with important compatibility or aim improvement problems.

Having outlined the above aim oriented theory of scientific method, I have now to confess that considerations to be developed below, arising from a discussion of the key notion of intelligibility, will lead us to an important modification of the above theory, at least as far as the theory applies to *physics*.

12. Intelligibility and the Search for Understanding. According to aim oriented empiricism, the fundamental goal of pure science is to improve our *understanding* of the world around us. We do not want merely to accumulate more and more truths about the world; nor do we want merely to predict successfully more and more phenomena with more and more accuracy. We want something much more than this, namely *understanding*. On this view, to say that a new theory makes a contribution to knowledge is to say just that the theory helps us to improve our understanding of the world around us.

It is, however, a main tenet of aim oriented empiricism that understanding is not to be had in *any* possible universe: it is only in a universe that possesses a certain inherent coherence, simplicity, harmony, unity, order, lawfulness or, as we may say, *intelligibility*, that that which we seek in seeking understanding exists to be found at all. On this view, to “understand” is just to see that ostensibly random, complex, arbitrary, diverse phenomena are really simple, coherent, unified, homogeneous or “intelligible.” Thus, our decision to pursue the scientific quest for understanding in effect commits us to taking entirely seriously the wholly speculative metaphysical idea that the universe *is* intelligible. It is as if our whole science is an attempt to discover more and more about a certain *object*—namely the inherent intelligibility of the universe, the intelligible *structure* of the universe. We do not have good reasons for holding that this object really does exist; we do not, in other words, have good reasons for holding that it is *true* that the universe is intelligible. But so passionately do we desire to discover intelligibility, of such supreme *value* do we hold the discovery of intelligibility to be, that, quite rationally, we are willing to

undertake our search even though we cannot be sure that that which we seek really does exist, and even though we cannot be sure that we have *found* that which we seek, however amazingly successful our scientific search may *appear* to have been.

According to aim oriented empiricism, then, scientific progress is not to be equated merely with developing theories which successfully predict more and more phenomena more and more accurately. In addition to this we require that our theories are of increasing inherent intelligibility. In fact, for real progress in science our very standards of intelligibility ought steadily to improve. That is, for real progress we should require not only that our theories are increasingly intelligible in some fixed sense of 'intelligible' but, in addition, we should require that our theories are increasingly intelligible, even though our very standards of intelligibility are themselves improving, becoming more powerful and exacting. Our ideals of intelligibility, enshrined in our chosen blueprint, ought themselves to evolve and improve. The greater the intelligibility of our best theories, the more our understanding of the world is thereby enhanced, the greater the contribution to knowledge, and the more perfectly the fundamental aims of pure science are realized. Understanding is the supreme goal, the supreme desideratum, of intellectual inquiry: it is thus ultimately in order that our interests shall be served, our intellectual ends realized, that we demand that our theories become more and more intelligible as science proceeds.

But what, it may be asked, *is* intelligibility (simplicity, coherence, unity, beauty, etc.)? What is it that we *mean*, at the broadest level, when we say that science seeks to discover intelligibility inherent in phenomena? Very roughly, I have indicated the kind of thing that it means to say that the universe is "intelligible"; it is to say that a metaphysical theory, more or less like one of those indicated in sections 6 and 11 above, is true. But what is it that these blueprints have in common that prompts us to declare that each blueprint provides us with a more or less "intelligible" picture of the universe? How, for example, are blueprints to be compared with respect to their "intelligibility"? The whole import of the argument so far has been to show that it is absolutely essential that we can make an *a priori* choice between rival blueprints with respect to their intelligibility if science is to be rational. We need to be able to make such *a priori* choices between rival possible metaphysical blueprints both from the standpoint of developing new scientific theories in a rational manner, and, perhaps even more important, from the point of view of choosing rationally between rival empirically successful scientific theories. Without *Rule 1* of section 11, in other words, the whole rationality of science founders. But if we are to choose between rival blueprints with respect to their intelligibility in a rational manner, we clearly must have some extremely broad, agreed idea as to what intelligibility is which is not to be explicated in terms of some *specific* blueprint. It is to this question of what 'intelligibility' ought to be interpreted to mean in general (at least as far as physics is concerned) that I now turn.

The first point that I wish to emphasize is that as long as we leave things sufficiently vague and intuitive, there can be no doubt at all that there is a meaningful notion of "intelligibility," not tied down to any *specific* blueprint, which plays an enormously important role in science. That science persistently seeks simplicity,

unity, coherence, order, lawfulness, harmony, beauty, explanatoriness or intelligibility in *some* sense of these terms, can hardly be doubted by anyone, whatever their views may be about the nature of scientific inquiry and scientific knowledge. Even the most casual glance at the history of physics, let us say, reveals the enormous importance that notions such as these have played in influencing the direction of research, and in influencing evaluations of contributions to knowledge. For example, the hope that nature may ultimately turn out to be simple, coherent, unified or intelligible can be seen to lie behind the seventeenth century idea that all natural phenomena arise as a consequence of the arrangements and motions of a few different sorts of elementary corpuscles, interacting only by contact; the same hope lies behind the Faraday-Einstein idea that everything is made up of one unified field; and it lies behind Wheeler's idea [18] that there is in the end only curved empty space with changing topological features. The same goal of seeking underlying simplicity, unity, beauty, intelligibility thus lies behind the advocacy of wildly different blueprints.

Some of the greatest contributions to science are precisely contributions which "unify" apparently diverse phenomena (often against a background of entirely different metaphysical blueprints): there is, for example, Newton's unification of the motion of terrestrial and astronomical bodies by means of his laws of motion and law of gravitation; Maxwell's unification of electricity, magnetism and optics (further unified by the special theory of relativity); Einstein's unification of gravitation and geometry. Again, there is the discovery of the relatively few chemical elements, from which all the millions of diverse chemical compounds can be built up; there is Mendeleev's discovery of a pattern in the properties of the elements; and the twentieth century discovery that all matter is built up of just three types of particles—electrons, protons and neutrons. There is the tremendous unification achieved by quantum theory—a few relatively simple physical postulates sufficing in principle to predict and explain all the vast diversity of physical and chemical properties of ordinary matter. And there is the discovery that all natural phenomena can be understood in terms of just four (or even possibly only three) kinds of forces or interactions. The whole program of physical inquiry is clearly committed to the discovery of unity, simplicity, order, coherence, harmony, symmetry (in some sense of these terms) at a fundamental level, and when this is not found—as for example in the case of the discovery of more and more so-called fundamental particles—the assumption is invariably made that a more basic kind of unity, simplicity, etc. underlies the ostensible complexity and multiplicity so far revealed. Even in the scientific search for natural law we must see the search for simplicity or coherence since, as I argued in **Part I**, footnote 4, the very notion of physical or natural law becomes empty of content unless we stipulate that a law must have a certain inherent simplicity, coherence or *nonaberrance*. And of course, finally, the fact that scientists invariably ignore, in an entirely *a priori* fashion, an infinity of *aberrant* theories, just as empirically successful as, or even more empirically successful than, the best chosen theory, indicates dramatically the all pervasive influence of simplicity considerations in science.

No one, then, can deny that the persistent search for simplicity, unity, coherence,

intelligibility, etc., in *some* sense of these terms constitutes an enormously important part of scientific inquiry. Everyone must agree that it is both *meaningful* and *true* to say that science seeks simplicity, unity, etc., in some sense of these terms. To confess that one has no idea at all what these terms may mean in the context of scientific inquiry is simply to confess one's ignorance of the realities of the scientific enterprise.

The task that lies before us, then, is not to define 'intelligibility' from scratch, as it were, as if we had before us an entirely uninterpreted term to which some kind of meaning had to be assigned out of the blue. Rather, the task is to uncover and solve the *problems* that our intuitive notion of intelligibility—unquestionably of supreme importance for science—runs into. We need to articulate and overcome the *objections* that appear to lie in the way of the truism: "It is both meaningful and true to say that science seeks to discover intelligibility inherent in phenomena."

Let us consider, then, why it is that so many scientists and philosophers of science find the whole idea of science seeking simplicity, unity, intelligibility, etc., highly problematic. Why is there quite general puzzlement as to what "simplicity" etc. can *mean* in a scientific context?

There are, I suggest, at least *four* reasons for this.

(i) In preferring simple, coherent, unifying, intelligible theories to complex, ugly, incoherent, unintelligible, *aberrant* theories in an entirely *a priori* fashion, we are, it seems, presupposing, in an *a priori* fashion, that the world itself is simple, coherent, intelligible etc. rather than complex, unintelligible, aberrant. But, according to orthodox views about science, it is the very essence of being "scientific" that one does not make permanent *a priori* presuppositions about the world. Acceptance of standard conceptions of scientific propriety thus render the thesis that science seeks intelligibility profoundly problematic. There appears to be no conceivable *rationale* for committing science to the search for intelligibility (as opposed to the search for truth, or for empirical growth *per se*).

(ii) Simplicity, coherence, unity, harmony, beauty, intelligibility appear to be essentially subjective, aesthetic features of theories. The idea that simplicity criteria play an important role in assessing the acceptability of theories in science thus appears to undermine the essential *objectivity* of science.

(iii) Some of the most important contributions to physics have consisted of theories widely judged to give a thoroughly *unintelligible* picture of the universe. In fact, it may almost seem that as physics has progressed, our view of the physical universe has become more and more strange and bizarre. There appears to be, then, a major objection to the idea that physics seeks *intelligibility*, progress in physics depending on whether or not our fundamental theories become more and more intelligible.

(iv) It is completely unclear how one is to distinguish between mere *terminological* simplicity or unity, which can always be procured in an entirely routine fashion, and is quite without significance for science, and genuine or—as we may call it—*semantic* simplicity, which is of real import for science.

Let us see whether these four problems can be resolved.

The first problem is easily solved: all we need do is reject standard empiricism and accept aim oriented empiricism in its stead!

As long as we retain some form of standard empiricism, there would appear to be no satisfactory solution to the problem. For according to standard empiricism, the fundamental aim of science is to discover more and more about the world *as such* (or about the phenomena *as such*) without making any permanent *a priori* presuppositions about the nature of the world (or the phenomena). On this view, it must be the very *essence* of being scientific that permanent *a priori* presuppositions about the world are not made. But, as we saw in **Part I**, to place a premium on simple, coherent, intelligible theories, to the extent of just ignoring empirically successful complex, incoherent, aberrant theories *is* in effect to presuppose, in a permanent *a priori* fashion, that Nature herself is simple, coherent, intelligible. Within the confines of standard empiricism there can be no satisfactory way of reconciling the search for truth *per se* with the search for intelligibility. The more weight we give to the idea that pure science is a passionate, singleminded search for explanations, for intelligibility, so, according to standard empiricism, the more irrational and unscientific science must become. Just that which constitutes science's greatest glory is for standard empiricism one of the more disreputable and problematic aspects of the scientific enterprise—an embarrassment to be "explained away."

As long as we retain the basic tenet of standard empiricism, the only way to lessen the severity of this fundamental problem is, in general terms, to hold that science seeks to explain, seeks to discover intelligibility, in the weakest, the least demanding, sense of these terms possible. The more nearly science forgoes the search for explanations, the more nearly science becomes merely an attempt to predict more and more phenomena *per se*, the more rational and scientific science will be, according to standard empiricism. And if we look at the main philosophical views about the nature of scientific inquiry and scientific knowledge, what we see is precisely a number of ingenious attempts to "explain away" the search for explanation, for intelligibility. (This is true even of Kant, since Kant entirely excludes the possibility that science might discover intelligibility inherent in the real, noumenal world.) The Popperian school, for example, seeks to explicate "simplicity" or "explanation" in terms of high empirical content ([14], pp. 136–145), or in terms of the capacity of a research program successfully to anticipate novel phenomena ([8], pp. 173–176). Conventionalism and logical empiricism seek to reduce the search for explanation to a search for convenient systematization of the empirical content of science which, it is held, is embodied entirely in the sum total of accepted empirical laws.

The moment we reject standard empiricism, and accept aim oriented empiricism in its stead, the whole problem situation is, however, dramatically changed. For according to aim oriented empiricism, the thesis that science seeks intelligibility does not threaten to undermine the rationality of science but, on the contrary, lies at the very heart of the rationality of science. The stronger, the more demanding, the more *important* the sense in which science seeks to discover intelligibility, so the

more *rational* science becomes. There is, then, an important sense in which the notions of "explanation," "simplicity," "unity," "intelligibility," etc., are indeed highly problematic if we accept standard empiricism, but cease to be problematic if we accept aim oriented empiricism.

To some extent, the second problem concerning the inevitable *subjectivity* of intelligibility is also to be resolved by rejecting standard empiricism and accepting aim oriented empiricism in its stead. As long as standard empiricism is accepted, as long as we hold that the fundamental aim of science is to discover truth *per se*, it is very difficult to see how such quasi-evaluative, aesthetic, emotive considerations as beauty, simplicity, harmony, can conceivably play any *rational* role in science at all. If our basic aim is to discover truth *per se* then all such subjective considerations ought, from a rational standpoint, to be ruthlessly excluded from science altogether.

If we adopt aim oriented empiricism, however, all this is quite different. According to aim oriented empiricism, if pure science is to be a rational enterprise it is absolutely essential that that which we seek in pursuing science should be something that we passionately hope to find, and hold to be of supreme intellectual value. According to aim oriented empiricism, a science denuded of all human feelings and values could not conceivably be rational. It is, in other words, absolutely essential to the rationality of science that science seeks to discover "beauty" or "harmony" in Nature in a sense which can move our hearts!

But this is not to say that objectivity must be abandoned. According to aim oriented empiricism objectivity is indeed a vital aspect of science. For in pursuing scientific inquiry, our concern is not merely to procure the experience, the hallucination one might say, of perceiving beauty or intelligibility in the world; our passionate concern is to discover intelligibility which really does exist inherent in the world. We are all too aware of how easy it is to have a very powerful and apparently wholly convincing *experience* of perceiving intelligibility in the world (a multitude of different ideas all apparently falling into place to form a coherent whole), arising from all kinds of fortuitous causes. Just because it is so easy to be deluded in this field, we subject our "experiences" of perceiving intelligibility to fierce interpersonal criticism and reappraisal. At the heart of science there is a kind of agonized scepticism concerning the authenticity of personal experiences of illumination, and out of this scepticism comes the essentially interpersonal, objective character of science. Thus according to aim oriented empiricism, the demand for objectivity in science should not lead us to *reject* the personal, the emotional and the evaluative altogether; rather the demand for objectivity arises out of a concern that we should, insofar as this is possible, discover intelligibility which does actually, objectively exist in the world, there being, as a result, a concern to subject our own personal, individual experiences of perceiving order and harmony in the world to objective, interpersonal criticisms and checks.

There is, however, a second and perhaps even more important interpretation to be put upon the demand for objectivity, from the standpoint of aim oriented empiricism. We may regard the demand for objectivity as itself an intelligibility requirement.

The point can be put like this. If there is one person—such as God—who has a

very special status in the universe, then it is understandable that the universe is intelligible in terms of that person's own personal standards of intelligibility. Likewise, if there is a very special group of people, with a special cosmic status, then it is understandable that the universe should be intelligible in terms of the group's particular culture and standards of intelligibility. But if we deny the existence of any such cosmically preferred people, then it becomes clear that it would be very odd indeed—in fact incomprehensible—if the universe were so structured that quite arbitrarily it only made sense to one particular group of people with their own rather idiosyncratic ideas about what is intelligible, and what is unintelligible. Once we acknowledge that there are no cosmically preferred people, it becomes clear that the universe can only be intelligible if it is at least potentially intelligible to *all* people, and not only intelligible to a particular group of people who happen to live during a particular epoch, with a particular culture, on a particular planet, with such and such a neurological make up.

Thus we see that the requirement that intelligibility criteria be objective is itself an *intelligibility requirement*. In certain respects intelligibility is a more fundamental feature of science than is objectivity. The demand for objectivity is to a considerable extent to be understood in terms of the more fundamental demand for intelligibility. This simple consideration shows that it is completely unrealistic to argue that the requirement that scientific knowledge must be objective must lead us to banish intelligibility as a basic requirement from science, on the grounds that intelligibility must be irredeemably subjective.

If we look at the actual development of human knowledge, one of the most moving and dramatic features of the whole story is the long painful way in which highly parochial, egocentric, anthropomorphic ideas of intelligibility have been slowly made a little less parochial, a little less egocentric and anthropomorphic. One sees this in the realization that the particular affairs of the tribe may not be of cosmic significance, and hence ought not legitimately to be used in order to explain cosmic occurrences. It is implicit in the idea that the earth might not after all be at the center of the cosmos. It lies behind the seventeenth century distinction between primary and secondary qualities, and the demand that explanations should be couched only in terms of the former. (Secondary qualities are of importance only to human beings, with their own highly specialized sensory apparatus and nervous system; beings with a different neurological make up could not become aware of "our" secondary qualities. Thus if a theory is to be objective—and hence *intelligible*—it ought not to make any reference to merely "human" secondary qualities.)

The search for increasing objectivity, and hence increasing intelligibility, can perhaps be seen to lie behind the Galilean idea that uniform motion is relative, not absolute, which denies the fundamental and egocentric Aristotelian idea that the earth is absolutely at rest, and thus constitutes a privileged reference frame. Einstein's bold generalization of the Galilean idea, embodied in his general principle of relativity, can be seen as a continuation of the same theme, in that Einstein's principle obviates the need to postulate a unique set of inertial frames of reference.

Thus the search for standards of intelligibility which are as unparochial as possible, which can be appreciated by the widest possible range of possible beings,

and not merely by human beings from a particular culture, is a fundamental feature of the development of knowledge.

The idea that objectivity is itself an intelligibility requirement has one extremely important consequence, namely this: the more intelligible things become, the more *unfamiliar* they ought to become. There is a widespread idea that to explain, to render intelligible, is to reduce the unfamiliar to the familiar. But the familiar is almost bound to be the parochial, the anthropomorphic, the culturally and socially egocentric; and all this means—once we grant that objectivity is an essential intelligibility requirement—that a theory which reduces the unfamiliar to the familiar will be highly unsatisfactory as a good explanatory theory. Implicit in the very *idea* of rendering things (objectively) intelligible or comprehensible is the idea that the more intelligible our theories become, so the stranger, the less familiar they will become. This has an important bearing on how we assess the progress of physics viewed as an attempt to discover intelligibility inherent in the world. If we adopt the widespread view that to explain, to render intelligible and comprehensible involves reducing the unfamiliar to the familiar, then we shall also adopt the widespread view that as physics has progressed things have become more and more unintelligible and inexplicable, our current theories being better explanatory theories only in the technical sense of being theories which successfully predict more phenomena. The view advocated here—namely that physics seeks increasing intelligibility—will seem bizarre in the extreme, and will appear to be decisively refuted by the actual development of physics. All this is, however, a mistake. It is entirely wrong to suppose that increasing intelligibility involves increasing familiarity. As we have seen, just the reverse of this is the case. A necessary condition for the universe to be (objectively) intelligible is that it should be very, very strange.

Recently, J. Wheeler was quoted as saying: “The universe is far stranger and more beautiful than we realize, and also far more simple. But we have no hope of seeing how simple it is until we first recognize how strange it is.” This remark sums up the case for aim articulation and aim improvement quite beautifully. Inherent in the very idea of the world being ultimately intelligible (beautiful, simple, harmonious, etc.), is the idea that the world is ultimately very strange (since that which is objectively intelligible will automatically be highly unfamiliar). Our automatic intuitions as to what is intelligible or comprehensible are thus not to be trusted, since these intuitions are almost bound to confuse intelligibility with its opposite, *familiarity*. We need to subject our ideas of intelligibility to lengthy rational criticism and objective appraisal, in order to strip away the merely familiar, and arrive at that which is both highly unfamiliar and truly intelligible. Here is perhaps the most powerful argument for the potential fruitfulness of aim articulation and aim improvement.

Let us now see whether we can make a little more precise and explicit what we ought to mean by “simplicity” or “intelligibility” etc., at least as far as these notions arise in the context of physics. Before we attempt to do this however, it is of decisive importance that we become quite clear as to what criteria we want a satisfactory theory of simplicity or intelligibility to fulfill—otherwise we shall spend

our time looking in the wrong direction, searching for something that we do not really want at all.

Here are seven desiderata that a theory of simplicity or intelligibility ought, I suggest, ideally to satisfy.

(a) A number of terms tend to be used in the context we are talking about, namely: 'simplicity', 'coherence', 'unity', 'orderliness', 'lawfulness', 'harmony', 'beauty', 'explanatoriness', 'intelligibility'. In **Part I** I suggested that there is an intuitive connection between these ideas. Ideally, however, we want one single clear intuitive idea, clearly at the heart of the aims of physics, from which all these other ideas can be seen to follow. If we have this, then we have available the possibility of improving and developing our ideas of simplicity, intelligibility, etc., in a rational fashion, as physics proceeds, since the fixed basic idea may provide the means for assessing the relative merits of different interpretations of 'simplicity', 'harmony', 'beauty', etc.

(b) The theory should satisfactorily explicate our intuitive judgments concerning simplicity etc., as these arise in the context of physical inquiry and the history of physics.

(c) The theory should satisfactorily explain why we judge the discovery of simplicity, intelligibility, etc., to be so supremely important, to have such supreme intellectual value. This is essential if we are to exhibit physics as a *rational* enterprise.

(d) The theory should enable us to distinguish clearly between mere terminological simplicity, which is relatively unimportant, and *semantic* simplicity, which is of genuine value for physics.

(e) The theory must fit in with the important requirement that intelligibility must be *objective*—and with the idea that objectivity is itself in part an intelligibility requirement.

(f) The theory must articulate a notion of intelligibility which satisfies the apparently hopelessly contradictory requirements of being both extremely flexible *and* extremely precise. Let me explain. To propose that physics should seek simplicity or intelligibility or understanding in some more or less definite sense of these terms is, as we have seen, to propose that physics should have a more or less definite, fixed, permanent aim or metaphysical blueprint—a blueprint for the construction of all possible, more specific blueprints for physics, as it were. We have here, then, something like the fulfillment of the rationalist's dream: to propose that we ought to mean such and such by 'understanding' or 'intelligibility' as far as physics is concerned is, inevitably, at one and the same time, to propose that the Ultimate, Comprehensive, True Theory, that the whole of physics seeks is, roughly and crudely, of such and such a type. From the mere *analysis* of what we ought to mean by 'understanding' we arrive at our first rough, approximate glimpse of the Ultimate Truth!

This consideration underlines vividly that any proposal as to what 'intelligibility' ought to mean in the context of physical inquiry must be put forward in a highly *tentative* fashion, and, at all costs, must not be too precise, too restrictive. It is of course essential, if physics is to be rational, that we are able to change, or at least modify, our basic blueprint as research proceeds: it is thus essential that any

proposal as to what ought to be the permanent, unchanging blueprint of physics is extremely imprecise or flexible.

On the other hand, we do not want our notion of intelligibility to be so vague, so imprecise, that no definite rules arise which enable us to appraise the relative intelligibility of rival blueprints.

We thus have the following dilemma. On the one hand, if science is to be rational, it is essential that we have a relatively *precise* notion of intelligibility, permanent through all the revolutionary upheavals in physics, continuing through all blueprint changes, so that we can rationally assess rival blueprints in an *a priori* manner with respect to their intelligibility. But on the other hand, to propose a specific, precise, permanent notion of intelligibility for physics is, *ipso facto*, to propose a more or less precise, specific metaphysical blueprint for physics; and if physics is to be rational it is essential that we be able to *change* the basic blueprint of physics, so that we do *not* have a precise, fixed blueprint. In short, if physics is to be rational it is absolutely essential that we both *do*, and *do not*, seek intelligibility in a fixed, precise sense (which looks almost like a *reductio ad absurdum* of the thesis that physics can be rational!).

A satisfactory theory of physical intelligibility clearly must somehow resolve this central dilemma. It must somehow do justice to both of the ostensibly contradictory requirements: extreme flexibility and extreme precision. I suggest that the theory to be sketched here does succeed in satisfying this all important, central requirement.

(g) A satisfactory theory of simplicity or intelligibility ought to enable us to appraise not merely the relative intelligibility of rival blueprints, but also the relative simplicity, coherence or unity of rival physical *theories*.

At first sight it might seem that once the basic tenets of aim oriented empiricism are accepted, as outlined in sections 6 to 11 above, then a theory of simplicity for scientific *theories* is no longer required, at least in the sense of having *fixed, general* criteria of simplicity. For of course, according to aim oriented empiricism, ideals of simplicity or intelligibility (as far as theories are concerned) are at any stage enshrined in the chosen blueprint of that science. And if the science is to proceed rationally then it is essential that the chosen blueprint changes and hopefully improves as the science develops. And as the blueprint changes, so too will the appropriate level 2 simplicity or intelligibility rules of theory acceptance change as well. Thus, if science is to proceed rationally, it is essential that level 2 simplicity criteria (for theories) *change*, both from science to science at a given time, and within one science from one time to another. Traditional attempts to develop fixed, general criteria of simplicity for theories—such as that made by Goodman [5]—seem, then, to be seriously misguided, in that such attempts do not take into account the rational need for science to have changing, evolving standards of simplicity or intelligibility for theories.

However, even though we grant all this, it must also be conceded that as far as the natural sciences are concerned, at least, it seems to be the case that we prefer simple theories (in some sense of ‘simple’) in a way which is indifferent to, or overrides, changes of blueprint. Given a definite blueprint for physics, such as the

corpuscular hypothesis, or Einstein's unified field blueprint, and a list of corresponding level 2 intelligibility criteria, then it will, it seems, always be possible to develop an indefinite number of theories compatible with such intelligibility conditions, most of which would, however, be wholly unacceptable in an entirely *a priori* fashion in that the systems of equations involved would be incredibly *complex*. At once we are confronted by the traditional problem of distinguishing between genuine and merely terminological simplicity. An adequate theory of simplicity or intelligibility, put forward within the context of aim oriented empiricism, clearly must be able to resolve this problem.

Let us see, then, whether we can sketch a theory of intelligibility which satisfies the above seven requirements.

13. The Search for Understanding in Physics. A key intuitive idea behind the search for intelligibility or understanding in physics is, I suggest, that in seeking to understand we seek to discover why things are, in some sense, *necessarily* the way they are and not some other way. Given that some state of affairs changes in time, we wish to understand why that state of affairs changes in just this way, and not in some other way that we can imagine. Given that objects of some common type can be found to have different properties or features, we wish to understand why these objects have just *these* different properties or features, and not others that we can conceive of. We seek to "understand" change and diversity by attempting to discover some property or feature which is *invariant* through change, which is common to the diversity, and which can only exist if the change occurs in precisely the way in which it does. Or, in other words, we seek to discover a fixed *pattern* in events, a pattern which would be utterly disrupted if the events were to occur in any way other than the way in which they do. (A pattern is, of course, just something that is invariant under certain transformations or changes.) It is precisely by discovering such fixed patterns in phenomena, such invariant, persistent features through change and diversity, which cannot exist unless the changes occur in just the way in which they do, that we are able to see why it is that, in a sense, events could not occur in a way other than the way that they do, so that there is a kind of *necessity* to the way in which events unfold. In fact, given that there really does exist an invariant property of the required type, then, relative to this, the changes will be necessary in a strictly logical sense. (For a brief sketch of the essentialist conception of science that is in effect being appealed to here, see [9], reprinted in [16].)

Having discovered a *pattern* in a range of phenomena, at once we have, of course, the further problem of understanding why there should be just *this* pattern, this invariant feature, here, and not some other pattern, some other invariant feature. Here again, we seek to discover why things cannot be other than the way they are by attempting to discover a higher level common pattern encompassing ostensibly different patterns. We seek to discover how one known pattern can be transformed into other known patterns, the manner of such transformations being such, however, that there is a *further* pattern inherent in these changes of pattern, which would be destroyed if these changes were to occur in a way even slightly different from the way in which they do occur.

Consider for example the Galilean law of free fall. This in effect asserts that throughout an object's changes of position and changes of velocity during free fall (in a vacuum) there is something constant and invariant, namely the acceleration of the object: furthermore, if the object were to change its position and velocity in any way other than the way in which it does, this invariant feature would be destroyed. To this extent, then, this law enables us to understand why the object falls as it does. But of course the value of Galileo's law does not stop here: for the law asserts the invariance of acceleration through a vast number of changes, namely changes of locality, time, shape, weight, chemical composition, in addition to changes of height and velocity. The capacity to fall with constant acceleration thus persists through an immensely wide range of changes, and herein lies at least one reason for the enormous importance of the law, granted that our concern is to "understand" in the general sense being explicated here. (The fact that the law refers to a specific object—the earth—and is thus not a law at all in the conventional sense is a matter of relative unimportance when this is compared with the universality of the law in other respects.)

Given Galileo's law of free fall, however, it is natural to ask: Why is the acceleration of objects just 32 ft. per sec.², and not half or double this value, for example? And why is the *acceleration* constant, and not the velocity, or the change of acceleration with time? In order to answer these further problems of understanding, we need to discover what freely falling objects near the surface of the earth have in common with other moving objects, such as planets moving round the sun. Newton's laws, of course, provide us with far deeper understanding by specifying what is invariant through all moving objects everywhere—namely for any two objects of any masses M_1 and M_2 at any distance d apart, $F = GM_1M_2/d^2$. Galileo's pattern can (in principle) be changed in at least three ways: by changing the mass of the earth; by letting the object fall through more than an infinitely short distance; by varying nearby massive bodies other than the earth. Newton's law specifies the persistent pattern through these changes. The ultimate hope behind the Newtonian-Boscovichean blueprint was of course that *all* change, including all chemical changes, all differences of chemical and physical properties, could be understood in terms of one simple, universal pattern: invariant point-atoms change their relative positions and motions in a manner precisely determined by the common invariant properties of point-atoms, namely the force field surrounding each point-atom. According to the Newtonian-Boscovichean blueprint there is in reality only an inessential difference between any two events, states of affairs, substances, or phenomena in the world; all such differences are due to there being different numbers of point-atoms being involved (any two of which are essentially alike), differently arranged or in different kinds of relative motion. We can thus say that according to the Newtonian-Boscovichean blueprint, any event, sequence of events or substance is essentially similar to any other event, sequence of events or substance, and it is in this sense that this blueprint specifies one universal fixed pattern inherent in Nature.

Here, then, is what all the different blueprints referred to above have in common: each blueprint specifies in very general terms that which is invariant or unchanging

through change and diversity, and which is itself responsible for the apparent change and diversity. Each blueprint is to be interpreted as specifying in the most general terms how things are similar and dissimilar—even those blueprints which would ordinarily be held to be grossly “unscientific.” To assert that everything occurs in order to fulfill the will of God is in effect to specify one invariant feature that any two events have in common.

Given that the permanent and fundamental aim of physics is to discover a necessity, a lack of arbitrariness, in ostensibly diverse phenomena, in the sense indicated here, we can at once understand why *ad hocness* or *aberrance* is abhorred in physics, and why physics is committed to a program of overall theoretical unification. An *ad hoc* or aberrant theory is of course a theory which postulates a *flawed* pattern, a uniformity with some arbitrary exception: such a theory fails completely to specify anything that is invariant through the change or diversity with which the theory deals, and thus fails completely to explain why things cannot be otherwise than the way they are. Again, as long as theoretical physics postulates a number of *distinct* patterns (such as the distinct patterns of Maxwell and Newton, or quantum theory and general relativity today), there remains the problem of understanding why, given one pattern, the other *must* exist. If, however, we can discover a more comprehensive pattern which enables us to see what it is that remains invariant in moving from one to the other of the previously discovered patterns, then relative to the existence of the more comprehensive pattern, we can understand why the previously discovered patterns must be (at least approximately) as they are. Ideally we want the more comprehensive pattern to enable us to see how one previously discovered pattern can be continuously transformed (in some sense) so that it becomes an ostensibly quite different previously discovered pattern (rather in the manner in which Newtonian theory enables us to transform Galileo’s law so that it becomes one or other of Kepler’s laws).

According to this view, then, in seeking understanding physics presupposes that there is a comprehensive, overall pattern in Nature, a concealed homogeneity or uniformity invariant through all change and diversity, it not being logically possible for events to unfold in any way other than the way they do, granted that the postulated pattern really does exist. Nature is a work of art of great beauty, so much so that any one slight detail being a little different would utterly destroy the whole effect. This at least is the *ideal* that physical inquiry hopes to move towards discovering, contributions to knowledge ultimately being assessed in terms of the extent to which they help us move towards realizing this goal.

But now we run into a difficulty. Intuitively there appears to be a sharp distinction between an array of objects (events, sequences of events, properties, substances or phenomena) which on the one hand does exhibit a pattern, and on the other hand does not. If we imagine a number of samples of possible wall paper patterns, for example, it would seem quite clear that we would be able to distinguish unambiguously between those samples which exhibit a pattern, those that exhibit a flawed pattern, and those that are merely a horrible mess. Unfortunately, it seems that, ultimately, a pattern, a uniformity, is something that only has an existence relative to certain conventions concerning what is to count as “similar” and “dis-

similar.” Relative to an appropriate choice of conventions, an appropriate choice of what is to count as similar and dissimilar, any kind of chaos can be transformed into a beautifully regular pattern exhibiting wonderful homogeneity and uniformity, and *vice versa*. Thus, although at first sight one is inclined to suppose that the assertion that there is one fixed comprehensive pattern in all phenomena amounts to a substantial assertion, it would seem that the assertion is in fact empty of content. At first sight it appears entirely meaningful to draw a sharp and completely general distinction between a perfect pattern, and a flawed pattern (a distinction absolutely essential to the notion of “necessity” or “nonarbitrariness” being explicated here). It would seem, however, that no such general distinction can be drawn, since an array of events or phenomena that exhibits a flawed pattern, from one point of view, can always be transformed into a perfect and beautiful pattern, from a somewhat different point of view. Or, in other words, an *aberrant* theory, which is beautifully simple apart from one entirely *ad hoc*, arbitrary “flaw,” can, it seems, always be transformed into a simple and nonaberrant theory by the introduction of new terminology, corresponding to new conventions as to what is to count as similar and dissimilar.

One might try to overcome this objection by developing a purely formal, abstract theory of patterns which somehow succeeds in making a completely general, abstract distinction between “real” patterns and “pseudo” patterns. I do not believe, however, that such a thing can be done. The notion of “pattern,” of something being “invariant” through change or diversity, only becomes meaningful, I believe, relative to more or less specific conventions concerning similarity and dissimilarity, however general and abstract these conventions may be.

Thus it seems that in order to make sense of the thesis that there is a fixed universal “pattern” inherent in phenomena (the world being ultimately “intelligible”) we need to specify the *kind* of way in which things are ultimately similar and dissimilar. But how can we do this? If we grant that the universe is ultimately intelligible (in the intuitive sense we are attempting to explicate) then it is entirely reasonable to suppose that we do *not* at present know how things ultimately are similar and dissimilar—with respect to which all phenomena can in principle be seen to fall into a beautifully simple pattern. Even such basic ideas as that the geometry of space is Euclidean, or that the geometry of space-time is Riemannian, or that space-time is continuous and simply connected, are suspect.

Clearly the kind of thing we need to do in order to overcome this problem is to specify some particular conventions, *C*, as to how things are ultimately similar and dissimilar, and then specify a set of rules, *R*, which lay down certain restrictions on the ways in which *C* can be modified, so that *C* and *R* together in effect specify a restricted set of conventions. The assertion that the universe is intelligible then becomes: there exist conventions *C'* which can be arrived at from *C* by applying the rules *R*, all phenomena occurring in accordance with some beautifully simple pattern with respect to the conventions *C'*. In this way we can hope to articulate a notion of “pattern” (and thus of “intelligibility”) which can be formulated in terms of the concepts and theories available to us today, a notion which is, nevertheless, sufficiently flexible to allow for fundamental conceptual and theoretical innovation,

without thereby becoming so flexible as to be entirely meaningless. This kind of procedure for explicating a notion of “intelligibility” clearly leaves room for immense flexibility. At one extreme, we might interpret the rules R in such a way that any group C' of conventions is legitimate as long as it is possible for us to specify these conventions, in a finite amount of time, given our present methods for introducing and specifying new ideas, new concepts, new conventions (whatever precisely these methods may be). As long as we allow that we can meaningfully and consistently refer to possible conventions that we cannot possibly in fact “specify,” “formulate” or “construct,” then the above loosest possible interpretation of intelligibility will still be extremely restrictive in that there will be an infinity of possible universes that will be *unintelligible*. It would seem desirable, however, to explicate a notion of intelligibility that is far, far more restrictive than this extreme of flexibility. It would seem reasonable to demand that for the universe to be “intelligible” it is necessary for it to be intelligible with respect to conventions concerning similarity and dissimilarity which are in some way a generalization and objectivization of the conventions we uphold now. In other words, given that we interpret present day possible blueprints for physics as in effect specifying conventions of similarity and dissimilarity that we uphold at the moment, then, for the universe to be intelligible in an extremely restrictive sense, we require that there is some generalization and objectivization of present blueprint ideas with respect to which all phenomena fall into a beautifully simple “pattern.”

The claim that I wish to make is that the conception of “intelligibility” that I have very roughly indicated here gives every hope—when further developed and articulated—of satisfying all the desiderata spelled out above. In particular, this notion of intelligibility succeeds in meeting condition (f) above, in that it succeeds in meeting the apparently hopelessly contradictory requirements of being both extremely flexible and extremely precise. The all important point here is this. We may have considerable freedom in considering various possible candidates for similarity and dissimilarity conventions, C_1 , C_2 , C_3 , etc. But if a given set of conventions C_n is to be an acceptable candidate for the conceptual foundations of physics, highly exacting, fixed, *a priori* requirements must be satisfied. It must be the case that with respect to C_n , only relatively few, beautifully simple patterns can be formulated—so that relative to C_n there is a sharp distinction between perfect patterns on the one hand, and flawed patterns or chaos on the other hand. In other words, for C_n to be an acceptable possible conceptual foundation for physics, it must be the case that only relatively few comprehensive theories are possible which take on a dazzlingly simple form when formulated in terms of the conventions, or terminology, of C_n . The notion of intelligibility indicated here is thus highly flexible, in that it allows for fundamental conceptual change and revolution, and at the same time is highly restrictive, in that a fixed, extremely exacting requirement is placed upon any proposed fundamental conceptual scheme for physics. In addition, the notion of intelligibility indicated here enables us to understand, quite naturally, the enormously important role that “simplicity” plays in physics, and enables us, as we shall see, to differentiate between genuine simplicity and mere terminological simplicity.

We have arrived, then, at a position which gives a fundamentally new interpretation to the notion of “blueprint” in physics. So far, a blueprint has been quite simply a rather vague, imprecise sketch of the ideal future physical theory. For a physical theory T to be acceptable, relative to the acceptability of a blueprint M , we required simply that T be compatible with M — T , in effect, being a precise, testable version of M . This conception of the role of the “metaphysical blueprint” in physics then led us straight to the problem of what precisely it could *mean* to say that a blueprint M “enshrines” or “embodies” our ideals of intelligibility for physics—and in particular, the problem of how rival blueprints can be compared with respect to their intelligibility. The viewpoint that has been developed in this section adds a whole new dimension to the role of the “blueprint” in physical inquiry. According to our present position, a blueprint M , in specifying in completely general, comprehensive terms how things are similar and dissimilar, is to be interpreted as in effect specifying a set of conventions, C , with respect to which all natural phenomena can be seen to fall into a beautifully simple “pattern.” In other words, in advocating that M is the ideal blueprint for physics, we are advocating that if we develop terminology, t_M , specifically designed to be especially appropriate to M , so that the similarities and dissimilarities, the universal invariances, postulated by M , are mirrored in the terminology t_M , then a beautifully simple, comprehensive true theory T can be formulated in terms of t_M . The terminology t_M in effect embodies conventions concerning similarity and dissimilarity with respect to which the idea of “pattern” takes on a meaning. And for there to be a universal pattern in phenomena, in this sense, it is necessary that there should be a comprehensive true theory T which takes on an especially simple form when formulated in terms of t_M . Thus for a theory T to be acceptable, relative to a blueprint M , we require far, far more than merely that T should be compatible with M , or should be a precise, testable version of M . We require that T should attribute a simple pattern to phenomena, when formulated in terms of terminology especially appropriate to M . And a blueprint M embodies ideals of intelligibility in physics precisely *via* this extremely restrictive relationship that a theory must have to M , to be acceptable.

14. The Simplicity of Theories. The basic idea of the theory of simplicity hinted at in the last paragraph is that genuine or *semantic* simplicity is a property, not of a theory T on its own, but rather of a theory T *relative to a blueprint* M . As a first approximation, we may explicate the notion of semantic simplicity like this. We have, let us suppose, a blueprint M , and a number of interpreted languages or terminologies L_1, \dots, L_n , each of which is more or less “especially appropriate” (in a sense to be discussed below) for the formulation of the main tenets of M . For T to have semantic simplicity (relative to M) we require that T takes on an especially simple form, i.e. has *terminological* simplicity, when formulated in one or other of L_1, \dots, L_n . Or, as a slight refinement, for T to have semantic simplicity we require that T becomes more and more terminologically simple as it is formulated in languages more and more “especially appropriate” to M . By ‘terminological simplicity’, here, we may mean something quite crude such as that the fewer the number of symbols in an expression so the greater the terminological simplicity of

the expression (though no doubt this crude notion can be made more sophisticated).

The semantic notion of simplicity to be advocated here is thus explicated *via* the purely syntactical notion of terminological simplicity. The semantic notion is, however, itself a nonsyntactical notion. Semantic simplicity is a property of a pair of *propositions* T and M (or a property of one proposition T relative to a second M) entirely independent of all questions of what language or terminology T and M are formulated in. The question of whether or not T has semantic simplicity, relative to M , concerns the question of whether it is *possible* to have a language “especially appropriate” to M in terms of which T may be formulated with terminological simplicity; and the answer we receive to this question is entirely unaffected by what language we formulate T and M in when *asking* the question.

Essentially the same point may be put in a slightly different fashion, as follows. The fact that the theory being advocated here explicates semantic simplicity *via* the purely syntactical notion of terminological simplicity does not constitute a defect of the theory for the following reason. The basic *problem* that a theory of simplicity has to solve is to specify the circumstances under which terminological simplicity is an indication of *genuine* simplicity, and the circumstances under which it is nothing of the kind (terminological simplification being *merely* terminological). It is precisely this problem that the theory being proposed here succeeds in solving. If a theory T becomes more and more terminologically simple as a result of being formulated in languages more and more “especially appropriate” to an acceptable blueprint M , then this terminological simplification represents genuine or *semantic* simplification. Otherwise such simplification is *merely* terminological (and in all probability is of no significance for science).

Terminological simplification can always be achieved. However complex and intricate a theory may be, by introducing appropriate abbreviations, it can always be reduced to an ideally simple form, such as ‘ $A = O$ ’. Semantic simplification, however, cannot so lightly be achieved. Given any theory, T , however complex, then, if we are free to choose an appropriate M , we can no doubt always construct a language especially appropriate to (an appropriately chosen) M , so that T comes out as ‘ $A = O$ ’. But if we are presented with a specific T and a specific M , so that we are no longer free to choose M as we please, then our hands are more or less tied; it will not, in general, be the case that T takes on a simpler and simpler form as it is formulated in languages more and more appropriate to M .

Let us now look at the all important question of what it means to say that a language or terminology is “especially appropriate” to a blueprint M , or that languages L_1, L_2, \dots are increasingly “appropriate” to M . Roughly speaking, we need to find some way of picking out those languages which might be invented to formulate the basic tenets of M in complete ignorance of any specific theory T , so that no special terms are introduced designed specifically to reduce a complex T to an artificially simple form.

A blueprint M can be regarded as specifying certain sorts of entities with certain sorts of properties. Thus the Boscovichean blueprint postulates three dimensional Euclidean space, one dimensional continuous time, and point-atoms which interact by means of centrally directed forces which vary in a continuous fashion with

distance. The blueprint of general relativity postulates the four dimensional continuum of space-time, gravitation being a geometrical feature of this continuum, produced by mass-energy.

A language or terminology especially appropriate to a blueprint M clearly must include the means to refer to the sorts of entities specified by M , and the means to attribute the specified sorts of properties to these entities. One important restriction can however be placed on a language L that is to be especially appropriate to M : no extralogical terminology can be introduced which is such that it is possible for this terminology not to refer to anything, *even though M is true*. Thus, given the corpuscular hypothesis, we can introduce into a language L especially appropriate to this blueprint the terminological means to describe corpuscles of *any shape whatsoever*: we cannot, however, introduce terms that allow us to refer to corpuscles of certain restricted shapes only. (It is assumed here that the corpuscular blueprint places no restrictions on the shapes of corpuscles.)

In setting up languages more or less appropriate to a given blueprint we can proceed in stages. At one extreme we may develop a language L whose basic terms have meanings which are such that the terms could refer to entities of a far wider class than those postulated by the blueprint M . At the other extreme, the meanings of the terms of L may be much more tightly restricted so that they refer more or less exclusively to entities postulated by M . Thus, given the Boscovichean blueprint, our basic spatial terms may not incorporate *in their very meaning* the idea that space is Euclidean—although of course the language would incorporate the means to assert that space is Euclidean. A language “more appropriate” to the Boscovichean blueprint, however, would be a language whose basic spatial terms *do* incorporate in their meaning the idea that space is Euclidean, so that in terms of this language, the postulates of Euclidean geometry (apart from *existence* postulates) are *analytic*, and the spatial terms of the language only apply to physical space as long as space *is* Euclidean. Again, granted that the Boscovichean blueprint includes the postulate that mass is conserved, we have a choice. On the one hand, in order to refer to mass, we may introduce a three place predicate $m(x, a, t)$, taking real numbers, point-atoms and times as arguments, and then add the *contingent* postulate that the masses of point-atoms remain constant throughout time. On the other hand we may introduce a two place predicate $m(x, a)$, especially tailored to the ontology of the Boscovichean blueprint, it being implicit in the very meaning of the term ‘ $m(,)$ ’ that mass is conserved (so that conservation of mass is *analytic*). The more a language is “especially tailored” to the ontology of a blueprint in the above sense, so the more “especially appropriate” we may hold that language to be to the blueprint.

There is another, closely related way in which we may explicate the idea of a series of languages L_1, L_2, \dots being more and more “appropriate” to a blueprint. The fundamental function of a blueprint, it will be remembered, is to specify that which is invariant through all change and diversity: the more precisely a terminology itself “mirrors” the invariance postulated by a blueprint, so the more “appropriate” that language may be held to be to that blueprint. In other words, a blueprint M will give rise to a number of level 2 invariance principles $P_1, \dots P_n$, which assert

that nothing *essential* is changed by such and such (inessential) changes. A terminology especially appropriate to M will be a terminology which has built into it the invariance principles P_1, \dots, P_n , so that any specific *theory* formulated in terms of this terminology, will be such that it conforms to P_1, \dots, P_n . For example, the Boscovichean blueprint asserts that space is homogeneous and isotropic, so that a change of orientation, position or velocity with respect to space itself, can make no physical difference whatsoever. A terminology especially appropriate to the Boscovichean blueprint will be such that theories formulated in terms of this terminology will have their form entirely unaffected by changes of position, orientation and velocity of physical systems *with respect to space itself*. In other words, for a language L to be especially appropriate to a blueprint M , we require the following: If M asserts that nothing *essential* changes through the transformations $E_1 \rightarrow E_2 \rightarrow E_3, \dots$, then a description of E_1, E_2, E_3, \dots in terms of L should be such that these *descriptions* are all *essentially* the same (*terminologically* the same, that is). The Boscovichean blueprint asserts that stable systems of point-atoms—different “molecules”—can be ordered with respect to number of point-atoms involved, nothing *essential* changing from one possible molecule to the next, except for the addition of one point-atom. A language especially appropriate to the Boscovichean blueprint ought to mirror, terminologically, this important invariance. Again, to give a quite different example, once we accept the basic blueprint of general relativity, according to which space-time is Riemannian, gravitation being a geometrical feature of space-time, then we are committed to the idea that nothing *essential* is changed in moving from one set of continuous coordinate systems to another. A terminology that is especially appropriate to the blueprint of general relativity must capture this all important idea, and it is precisely this that the tensor calculus is designed to do.

At first sight it might seem that I have given two quite different explications of the notion of languages L_1, L_2, \dots being “more and more appropriate” to M . For the first explication it is required that L_1, L_2, \dots be successively more and more precisely tailored to the ontology of M ; for the second, it is required that L_1, L_2, \dots have built into them more and more invariance principles, corresponding to M . I suggest however that these two explications are not really all that different. The entities postulated by M are entities whose basic properties do not change; in developing a language whose basic terms reflect more and more precisely the ontology of M , the nature of the basic properties attributed to the entities postulated by M , one is automatically, at one and the same time, building into the language more and more of the invariance principles corresponding to M .

The circumstances under which terminological simplification indicates genuine or semantic simplification can, then, be specified like this. We have, let us suppose, a definite blueprint M , and a succession of languages or terminologies L_1, L_2, \dots successively more and more appropriate to M . (We assume that L_1, L_2, \dots are all such that the basic postulates of M get formulated in each language with maximum possible terminological simplicity.) If now a theory T takes on a simpler and simpler terminological form when formulated successively in L_1, L_2, \dots then, to that extent T has semantic simplicity, relative to M . And if, given two different theories T_1 and

T_2 , T_1 takes on successively a simpler and simpler form when formulated in languages L_1, L_2, \dots , whereas T_2 becomes increasingly terminologically complex, then T_1 is semantically more simple than T_2 , relative to M . Simplifying a theory has semantic or scientific import when it is achieved as a result of rewriting the theory in terminology more especially appropriate to an acceptable blueprint, M ; otherwise such simplification is without semantic import.

It may happen, of course, that given some M and T then if we choose languages L_1, L_2, \dots more and more appropriate to M , T takes on a dazzlingly simple form, whereas if we were to choose somewhat different, but equally legitimate languages L_1^*, L_2^*, \dots , T would remain, or would become, extremely complex. In this case, T is not semantically simple relative to M . Even worse, it may happen that a blueprint is such that it is *always* possible to construct two or more different languages or terminologies especially appropriate to the blueprint, *all* theories invariably being complicated in at least one of these languages. This, as we shall see in the next section, is to be judged as a serious flaw in the blueprint, rendering its acceptance highly problematic.

Perhaps the most striking exemplification of the above notion of semantic simplicity is provided by Einstein's general theory of relativity. Once the background metaphysical blueprint of general relativity is conceded—that is, once we grant the principles of equivalence and covariance, the idea that gravitation is a geometrical feature of space-time, the geometry of space-time being Riemannian (so that the conceptual apparatus of Riemannian geometry may be incorporated into our language)—then the equations of general relativity can be expressed with extraordinary (terminological) simplicity. In fact, once we stipulate that the theory be generally covariant, that it involve only the metric tensor and various geometrical objects obtained from its components and derivatives of these, that the equations be of the lowest possible order, and that the stress energy tensor be the source of gravitational effects, the Einstein field equations follow uniquely—so beautifully does the theory harmonize with its background metaphysical blueprint ([1], p. xix). And yet, as Hoffmann amusingly notes: “If written out in full instead of in the compact tensor notation, . . . [Einstein's field equations] would fill a huge book with intricate symbols” ([6], p. 122). Hoffmann goes on:

It may seem ridiculous to talk about beauty and near miracle after implying that the equations are ugly and cumbersome. Let us therefore ask a question. How did Einstein manage to find the equations? Could he have guessed the various terms—hundreds of thousands of them, or in one form, millions, and all of them highly unpleasant? Impossible. Then how did he find them? That is where the beauty and near miracle come in. For the tensor calculus contained stringent rules. Einstein imposed, for physical reasons, a few almost trifling conditions that, for the most part, had the effect of a request for simplicity. And when, having done so, he sought ten tensor equations in which gravitation would be represented *solely* by the ten quantities $g_{\mu\nu}$, he found that his hands were tied. Because of his insistence on simplicity, the tensor calculus gave him no choice. The field equations were determined *uniquely*. In the

tensor notation these equations are compact. Their power and their utter naturalness in both form and content give them an indescribable beauty. Suppose someone had actually written them out in full, term by term. A single misprint in the bookful of terms, a $1/2$ omitted or a 3 written instead of a 2, and the equations would not satisfy the condition of general covariance. ([6], pp. 122–123)

Several points vital to the basic themes of this essay emerge from this case of general relativity. First, it illustrates with dazzling clarity the thesis that we can only expect the fundamental equations of nature to have a “simple” form when written in terms of terminology appropriate to the entities of the proper blueprint. If formulated in terminology appropriate to some incorrect blueprint, the fundamental laws of nature can be expected to have only an impossibly intricate, complex form. This makes clear how vital blueprint articulation is for the progress of science—in particular the progress of theoretical physics. As Einstein himself always emphasized, if we are bereft of general principles to guide us in our search for new theories we simply cannot hope to stumble upon good new theories. In fact we can go even further, and say that without the proper blueprint, and hence the proper terminology, we could never have any serious reason to suppose we had formulated the proper equations *even if, as a matter of fact we had achieved this!* Einstein’s field equations, written out only in the form of millions of terms, could never be held to be confirmed or corroborated since we could never be able to *refute* all the theories—of equal *a priori* plausibility in this form—which make predictions only slightly different from general relativity. If we are to make possible both the *discovery* of a fundamentally new theory, and its *verification*, it is essential that we hit on the appropriate terminology: but we can only hope to do this if we choose the correct *blueprint*, the proper level 2 intelligibility principles.

In the field of high energy physics today we see precisely a search for appropriate symmetry and invariance principles (level 2 intelligibility principles) which in effect specify certain sorts of entities with certain sorts of properties, the hope being that a terminology, invented to describe these sorts of entities and properties in an especially simple, appropriate manner will be a terminology peculiarly appropriate to fundamental particle interactions, so that in terms of this terminology it will eventually be possible to write down reasonably simple dynamic equations for fundamental particle interactions.

The theory of simplicity sketched here helps perhaps to explain in a rather general way, why pure mathematics, pursuing its own ends, has so often generated ideas and terminology later on found to be enormously fruitful, if not essential, for physics. A pure mathematician such as Riemann [15] can be seen as engaged in the task of blueprint articulation and the development of mathematical machinery especially appropriate to the new blueprint ideas. No wonder this work turns out to be of value for physics. Up until now, however, blueprint articulation by mathematicians has proceeded in a somewhat fortuitous, hit and miss fashion, as far as physics is concerned. We need a much more conscious, deliberate study of blueprint problems, devoted especially to blueprints judged to be of relevance to future

physical theories, so that new fundamental physical theories may be developed in a progressive, *rational* fashion.

One needs to search far and wide for a theory as beautiful as general relativity, with as explicitly articulated a background blueprint. However, my claim is that throughout the field of the natural sciences, in assessing the semantic simplicity of a law or theory, invariably there lurk background blueprint assumptions which have influenced the choice of appropriate terminology, and in terms of which the law or theory has terminological simplicity. The theory of simplicity sketched here applies, I suggest, quite generally to physical theory even though it may not always be clear precisely what blueprint lies behind a theory, due to the reluctance of scientists to articulate blueprints, out of the mistaken idea that to do so is to be metaphysical, irrational and unscientific. It may even happen that a given theory has two or more distinct blueprints—as in the case of reformulations of Newtonian mechanics in terms of least action principles. Such reformulations of a theory, in that they provide a new, possible blueprint for the theory, must be regarded as highly creative developments, from the present standpoint, possibly effecting future developments in physics. The Hamiltonian reformulation of Newtonian theory is a case in point, since this has had an important role in the development of quantum theory.

The theory of simplicity sketched here is, of course, not confined to fundamental theories in physics. In seeking to develop a new theory in physics which is not of a fundamental character, we will of course be faced with a number of constraints: first, obvious conservation, invariance and symmetry principles, such as conservation of momentum and energy; second, relevant accepted fundamental theories (such as quantum theory); third, some specific *model* we have chosen for the phenomena under consideration. All these constraints will function as blueprint considerations; automatically, there will be terminology more or less “especially appropriate” to these blueprint considerations, and our task will be to develop a theory which has a simple form when expressed in this terminology (and which leads to equations that can be solved, a most important additional requirement that philosophers of science tend entirely to ignore).

It is interesting to note that in just the kind of context in which the problem of simplicity has often traditionally been considered by philosophers of science—the context of drawing a smooth continuous curve through a number of points on a graph—inevitably there must lurk in the background certain crucial blueprint ideas (though this has not usually been noticed). The point is this. A scientist only expects to obtain a “smooth” curve for certain sets of parameters, plotted one against the other in a certain manner. Merely in setting up a graph in one way rather than another, scientists are in effect accepting certain rather general “blueprint” ideas, and are choosing a terminology appropriate to these ideas. The information recorded by the smoothest of curves could after all easily be translated into a completely wild graph, if one adopted different background concepts. And if one chose two parameters at random to plot against each other one would *not* in general expect to obtain a smooth curve. Smooth curves arise relative to the choice of definite blueprint ideas, and appropriate terminology. In general terms, the

notion of simplicity sketched here seems to apply rather well to this classical “curve-fitting” situation.

15. Rules for the Appraisal of Intelligibility of Blueprints in Physics. At this point we really need an illustrative exercise in blueprint articulation. Were there space, I would like to provide such an exercise by showing how one can take the corpuscular blueprint as one’s starting point and then, by considering purely *a priori* intelligibility problems inherent in this blueprint, develop in turn Boscovich’s point-atom blueprint and Einstein’s unified field blueprint, in a progressive, rational fashion. In addition I would like to show how Einstein may be seen as developing both the special and general theories of relativity largely from a consideration of *a priori* intelligibility blueprint problems. In this way I would hope to illustrate and reinforce the following three cardinal points. First, the intelligibility of blueprints is a topic that can be rationally discussed (without there necessarily being instant decisive agreement on all issues) even in the absence of fully articulated, agreed rules for intelligibility assessment. Second, rational modification and improvement of a blueprint can lead to the development of profoundly novel ideas. Third, a self-conscious, articulate inquiry into intelligibility problems inherent in possible blueprints for physics may have enormously fruitful results in that in this way we may be able to develop entirely new fundamental physical theories in a rational and progressive fashion. Indeed it would seem (bearing in mind in particular the results of sections 13 and 14) that the blueprint articulation and improvement must constitute an essential preliminary to developing fundamentally new theories: for if our present blueprint ideas are false, as we can be reasonably sure they are, then fundamentally new physical theories will presumably receive only an impossibly complex form when written in terminology appropriate to our present (false) blueprint ideas. It is reasonable to suppose that the new fundamental theories we hope to discover in physics will only be recognizable as theories at all when we have developed new appropriate blueprint ideas. All in all, granted that our concern is to improve our understanding of the universe in as rational and successful a fashion as possible, the main task that confronts us today is to articulate a reasonably satisfactory unified blueprint for physics, and instigate a conscious tradition of discussion of intelligibility problems of blueprints, as an integral part of physical inquiry itself.

Unfortunately, there is not space here for an illustrative exercise in blueprint articulation. I must rest content with tentatively proposing the following five rules for assessing the intelligibility of blueprints (rules which in effect amplify *Rule 1* of section 11).

(1A) *Adequacy.* An acceptable blueprint must be such that it is at least conceivable that all relevant phenomena can be explained and understood in terms of the blueprint.

(1B) *Objectivity.* Given two blueprints B_1 and B_2 , if B_1 is intelligible in a more objective fashion than B_2 , then, other things being equal, B_1 is to be judged more intelligible than B_2 .

(1C) *Nonarbitrariness.* A blueprint which can only be made precise (to form a testable physical theory) in an arbitrary fashion, is to be judged to have an intel-

ligibility flaw on that account. Or, in other words, if a blueprint B is such that any precise version T of B has terminological simplicity only, and not semantic simplicity, then to that extent B lacks intelligibility. And on the other hand, a blueprint to which there corresponds a uniquely simple physical theory is to be judged highly intelligible, other things being equal.

(1D) *Generalization*. If B_1 is a generalization of B_2 , so that additional specific assumptions need to be added to B_1 to obtain B_2 , then, other things being equal, B_1 is a more intelligible blueprint than B_2 . In other words, the less specific, the less arbitrarily restricted, a blueprint is, so the more intelligible, and acceptable, it should be judged to be. (Generality is of course also desirable on purely epistemological grounds.)

(1E) *Coherence*. Suppose a blueprint consists of two parts, A plus B_1 , it being perfectly possible to construct an indefinite number of blueprints of the type $A + B_n$, where each B_n may differ radically from B_1 . (An example of this would be the Newtonian-Boscovichian blueprint, A , specifying the nature of space—three dimensional Euclidean geometry—and B_1 specifying the nature of matter, namely that matter is made up of point-atoms.) From the point of view developed in section 13, $A + B_1$ is unsatisfactory on intelligibility grounds in that there is no reason why, given A , B_1 should be true rather than B_2 or B_3 , etc. Suppose, however, we can generalize A to form A^* (so that A is just a special case of A^*) in such a fashion that B_1 can be dispensed with without loss of adequacy. In this case A^* is to be judged a very much more intelligible blueprint than $A + B_1$ in that it is a more coherent or unified blueprint. One way of assessing the importance of the blueprint changes brought about by general relativity is to argue that these changes take us some way towards the ideal of dispensing with matter altogether, and regarding matter as an aspect of space—a blueprint ideal that has, of course, been advocated and investigated by Wheeler, [18].

There is one point that may perhaps have troubled the reader. Is a blueprint, as explicated here, to be understood simply as a rather general untestable *theory*; or is it to be understood as a conceptual scheme or framework, embodying *conventions* of similarity and dissimilarity? Ambivalence on this point is deliberate. A blueprint begins its life, it would seem, as an attempt to specify concepts which any future physical theory must conform to. There is, initially, no idea of how an acceptable physical theory could possibly be set up outside the specified conceptual framework. The blueprint specifies, in other words, the limits of what is conceivable, or imaginable, as far as physics is concerned. But then blueprint developments, and in particular *mathematical* developments, make clear that from a conceptual standpoint, the chosen blueprint is simply a rather arbitrary special case of something much more general. Three dimensional Euclidean geometry is a special case of n dimensional Euclidean geometry. Euclidean geometry is a special case of alternative possible geometries, classified either in accordance with Klein's *Erlanger* program, or in accordance with the concepts of Riemannian geometry. And so on. From the vantagepoint of the much more general conceptual schemes, the original blueprint has become merely a rather imprecise *theory*. A fundamental task of blueprint articulation, then, is to remove arbitrary restrictions on accepted blueprint

ideas, that have become apparent as a result of conceptual developments in mathematics, at the same time ensuring that the more generalized blueprint satisfies the all important requirement of specifying a few semantically simple theories (rule *IC*).

16. Implications for Science. The ideas developed in this essay have, I believe, important implications not only for our understanding of science, but also for scientific practice itself. Aim oriented empiricism implies that the most important problem that faces a science at any stage of its development is to articulate possible fundamental aims or blueprints for that science, and then choose rationally the best possible candidate. It is essential to do this if there is to be rational *acceptance* of theories and a rational development of new theories. Conventional ideas of scientific propriety, however, positively discourage all attempts to choose metaphysical blueprints in a rational fashion. According to standard empiricism, it is the very essence of being scientific that, in the end, all theories, all theses about the world, are chosen in the light of purely *empirical* considerations. To attempt to choose rationally between different *metaphysical* ideas, which cannot be appraised empirically, is to cease to do science. It is to take up idle metaphysical speculation. The idea that metaphysical ideas may actually be more “firmly established” in a science than empirically verified theories, to the extent that if there is a clash between the two it may well be a “well established” empirical theory which is rejected, and not a metaphysical theory, is complete anathema to conventional ideas concerning proper scientific practice. All in all, the widespread acceptance of some form or other of standard empiricism, the *institutionalization* of this view in present day scientific practice, does much to discourage just that which we most need to do if we are to pursue scientific inquiry in a rational and successful fashion, namely constantly to put the chosen aims of a science under overt, explicit rational criticism and review. Standard empiricism is, of course, a completely unworkable creed; any scientist who tried to pursue science in accordance with the dictates of standard empiricism in a completely honest fashion would come to an instant and total standstill. At best, scientists only pay lip service to standard empiricism. But even this hypocritical allegiance to standard empiricism still has a crippling effect on scientific inquiry, for it still prevents scientists from performing the vital task of aim articulation, because to do so would be to violate all too obviously and blatantly the norms of good scientific behavior embodied in standard empiricism. Upholding some version of standard empiricism, most scientists (quite logically) do not believe in the possibility of a rational method of scientific discovery: as a result, not surprisingly, little attempt is made to develop new theories in a consciously rational fashion. The researcher proceeds with his eyes shut.

For some sciences, the fact that explicit aim articulation is outlawed as scientifically disreputable does not matter too much. For there are sciences which have relatively stable, unproblematic blueprints, consisting of the well established theories of some more fundamental science. Chemistry, for example, might be said to have quantum mechanics as its blueprint. Here, the incapacity of scientists to discuss explicitly metaphysical ideas in a scientific context, does not much matter.

In the case of a fundamental science such as physics itself, however, the situation is very different. Here blueprint articulation is an absolutely essential task that needs to be performed if any fundamental progress is to be achieved—even if this task is performed in private, as it were, by one or two individuals away from the critical eyes and ears of colleagues.

As I have tried to indicate during the course of this essay, it is not the case that *all* physicists accept some version or other of standard empiricism. There is at least one dazzling exception, namely Einstein. In developing both the special and general theories of relativity, Einstein pursued the path of articulating the best, the most intelligible aims for physics, thus enabling him to formulate two vital level 2 intelligibility requirements, namely the restricted principle of relativity, and the principle of general covariance, each of which played a vital role in the development of fundamental new theories. Even Einstein's famous 1905 photon hypothesis paper, which did so much to assist the birth of quantum mechanics, can be seen as essentially an exercise in aim articulation, since one may see the fundamental aim of that paper to be that of revealing an unavoidable, deep flaw in the scheme of classical physics which must arise when continuous radiation interacts with discrete matter. But Einstein did not only *do* science in accordance with the kind of aim oriented empiricist viewpoint advocated here, with quite extraordinary success.

Furthermore, after 1915, Einstein repeatedly explicitly advocated that theoretical physics should be pursued in this kind of way (see Einstein [2], [3]). The fact that Einstein's profound and original contributions to physics were developed in accordance with the rational method of discovery of aim oriented empiricism ought, I believe, to count somewhat in favor of this viewpoint. (From my own point of view, this success of aim oriented empiricism was entirely unexpected. I developed the ideas expressed here in an attempt to solve the Humean rationality problem, at first convinced that no scientist would have so violently violated conventional standards of good scientific behavior as to have consciously practiced aim oriented empiricism. It then gradually began to dawn on me that Einstein had in fact both practiced and advocated a position very close to the one put forward here.)

The extraordinary success of Einstein's new way of doing physics was too striking to be altogether ignored. And in fact Einstein's procedure of developing general methodological principles, level 2 intelligibility principles, has had a considerable impact on physical inquiry. For it was this procedure of Einstein's which led to the explicit study of conservation, invariance and symmetry principles, of just the type considered by Einstein. (Wigner, who has done much to develop the study of invariance and symmetry principles in physics, draws attention to the pioneering work of Einstein in this respect: see [19], pp. 5–7, 13, 14–23.) Today, invariance and symmetry principles play an increasingly important role in physical inquiry, particularly, as one would expect, in those domains of physics, such as fundamental particle physics, where fundamental new theories are being sought. A recent textbook has, for example, this to say:

... in particle physics much of our present dynamical theory is incomplete or unsatisfactory and it is reasonable to hope that many of the ideas and models

will be discarded or radically modified by the end of the century. In contrast, the results derived from symmetry theory are believed to be of much more permanent value. They are not a substitute for detailed dynamics, but they provide constraints which apply to any dynamical theory and their predictions can be explored and tested long before the details of such a theory have been worked out. Taken collectively, they provide a framework into which any subsequent theory must fit and in terms of which one can classify and interpret the present mass of experimental data. Thus in particle physics symmetry theory offers much more than just a set, however elegant, of insights and short cuts. It forms much of the basis on which the whole subject is built. ([4], preface)

Again, Wigner has this to say:

As far as the physical sciences are concerned, the role of invariance principles does not seem to be near exhaustion. We still seem to be far from the “universal law of nature.” We seem to be far from it, if indeed it exists, and, to paraphrase Poincaré, the present picture of four or five different types of interactions, with widely divergent properties, is not such as to permit the human mind to rest contented. Hence, invariance principles, giving a structure to the laws of nature, can be expected to act as guides also in the future and to help us to refine and unify our knowledge of the inanimate world. ([19], p. 36)

However, despite all this, it is, I believe, important to stress that the Einsteinian approach to physics has not been fully exploited by contemporary physicists, essentially because the *rationale* behind Einstein’s way of doing physics has not been understood. If some version of standard empiricism is accepted, it is very difficult to understand the curious status that conservation, invariance and symmetry principles have in physics—partly *a priori*, partly empirical, in that although these principles place restrictions on what is to count as an acceptable, intelligible theory, nevertheless these principles may be rejected in the light of empirical discoveries (for example, conservation of parity). Aim oriented empiricism, on the other hand, actually *predicts*, from first principles, that a mature science ought to have changing, evolving level 2 intelligibility principles.

According to conventional views of scientific inquiry, there can be no rational way in which important new invariance or symmetry principles can be developed (which in turn may lead to the rational development of new dynamic theories); there can be no rational way in which one can decide to accept or reject a given level 2 intelligibility principle, apart from taking into consideration the empirical success of a theory which incorporates or fails to incorporate the principle. But it is of course the whole essence of aim oriented empiricism that there is a rational (even though, of course, fallible and nonmechanical) way in which powerful new invariance principles, level 2 intelligibility principles, can be developed and assessed. In order to do this we need to articulate a coherent, unified blueprint for physics, which may be regarded as a harmonious collection of fundamental invariance principles formulated as a metaphysical theory about the nature of the universe;

we then seek to resolve awkward intelligibility problems inherent in this blueprint, in this way, in an entirely rational fashion developing what may well be a radically new blueprint which leads to radically new intelligibility principles and hopefully to radically new fundamental physical theories which achieve empirical success. It is only against the background of blueprint articulation and improvement that we can be in a position to develop new intelligibility principles, and choose between rival intelligibility principles in a rational fashion. Only in this way can we build into our methodology the all important idea that it is unity, coherence and generality that we seek at the most fundamental level of all. Today there is hardly an articulate, conscious tradition of blueprint improvement in this sense. And this is essentially because the reigning standards of scientific propriety render rational aim articulation and improvement an impossible enterprise, a foolish, unscientific exercise in metaphysical speculation. That which most needs to be done if we are to proceed rationally becomes that which is furthest removed from respectable scientific inquiry, once standard empiricism is accepted.

There are of course noble exceptions! Wheeler ([17], [18]) and Penrose [13], have, for example, sought to articulate comprehensive, coherent blueprints for a future physics. Such work is, however, generally judged to be essentially speculative, only being amenable to scientific assessment if empirical predictions are forthcoming. It is not appreciated that rational choice of scientific theory in terms of empirical considerations actually presupposes that we can choose rationally, in a more or less *a priori* fashion, between rival blueprints. Rational blueprint choice is essential if science is not to drown in an ocean of empirically successful aberrant theories.

The general neglect of the study of blueprint articulation in physical inquiry has led to the extraordinary situation that there is today no even half way acceptable unified blueprint for physics. (In this respect the situation is far worse than the situation in the seventeenth, eighteenth and nineteenth centuries, when relatively intelligible, plausible, unified blueprints for physical inquiry did exist.) It is not simply that the two fundamental theories of contemporary physics, namely quantum theory and general relativity, fail to fit together harmoniously: much more serious is the fact that the fundamental blueprint ideas behind these two theories are in violent collision. According to the aim oriented empiricist standpoint, then, the vital task for theoretical physics today is not the technical one of uniting quantum theory and general relativity by "quantizing" general relativity perhaps: rather the task is to articulate the blueprint ideas behind the two theories, and work out how these blueprint ideas can be modified so as to form a more general, nonarbitrary, conceptual unity. Only then can we be in a position seriously to contemplate developing a new unified fundamental theory.

Very tentatively, I would suggest that the fundamental problem that needs to be solved if an even remotely adequate aim or blueprint for modern physics is to be developed, can be put like this: How can one weld together the fundamentally *indeterministic* aspect of things that seems to have been irreversibly ushered in with quantum mechanics with the kind of "classical," microrealistic, field theoretic viewpoint advocated by Einstein?

Briefly, let me try to indicate this problem as I see it. One very striking feature of

quantum mechanics (QM) is that it is a theory which makes predictions exclusively about the outcome of performing measurements on microsystems. QM does not make predictions about the states of microsystems *as such*; rather the theory only makes purely *conditional* predictions about what *would* be found *if* a measurement were made.

This basic feature of QM ensures that QM must rely on classical physics, or on some additional theory, for a treatment of the measuring process itself. QM alone cannot issue in any *unconditional* predictions whatsoever; it is only QM *plus* some relevant part of classical physics which actually issues in physical predictions. And this in turn means that QM is an *aberrant* theory, in that the theory which results in physical predictions consists of the postulates of QM plus the purely *ad hoc* relevant part of classical physics which is applied to the measuring instrument.

In order to reformulate QM as a theory which can stand on its own feet, and which does not need to presuppose, in a wholly *ad hoc* way, some part of classical physics, it is clearly necessary to develop a version of QM which is *microrealistic*, which provides a consistent model for microphenomena. For as long as QM is a theory about microsystems interacting with *macrosystems* (measuring instruments) QM will need to rely on some purely *ad hoc* additional theory for a treatment of the macrosystems. Only if QM is developed as a theory about *microsystem interacting with microsystem* (macrosystems simply emerging as the outcome of quantum mechanical interactions between many microsystems) will QM become non-*ad hoc*, nonaberrant. Thus only a *microrealistic* version of QM can be ultimately nonaberrant and acceptable. (See [10], [11], and [12] for a fuller statement of this argument.)

At the moment, however, we do not have an idea of what a theory which is both indeterministic and microrealistic would look like, even in very general terms (see, however, [12]). It is the development of such a general idea that seems to me to be the most important task of aim articulation in physics today. We need to articulate in general terms possible kinds of fundamentally stochastic microrealistic theories, bearing in mind Einstein's important considerations which suggest that only field theories can be ultimately intelligible. Once we have a general idea of what an indeterministic microrealistic field theoretic theory looks like, we can then go on to consider in more detail how a *physical* theory of this type might be developed, and whether such a theory might help to resolve such outstanding fundamental problems as the reconciliation of QM and general relativity, and the problem of singularities in present day field theories.

The main point that I wish to emphasize however is this. According to the aim oriented empiricist viewpoint advocated here, the absence of any viable, coherent, unified aim or blueprint for modern physics indicates that theoretical physics today is in a state of *crisis*. Without such a coherent aim, we cannot have any clear idea of where we are going, of what lines of inquiry we ought to pursue. We do not even have an adequate basis for assessing the acceptability of physical theories already proposed. It was just this that Einstein knew so well, and articulated so clearly, particularly in relation to the ultimate unacceptability—or problematic character—of QM.

Nothing could illustrate more vividly how radically Einstein's views about the nature and aims of physics differed from the viewpoint upheld by his contemporaries than Einstein's break with his contemporaries over QM. For Einstein, QM was a profoundly unintelligible theory because it abandons microrealism, and fails to be compatible with the best blueprint for physics. Despite all its enormous empirical success, QM was, therefore, for Einstein, quite rationally, a profoundly *problematic* theory, offering no guidelines for future developments in physics. To Einstein's contemporaries, however, all this was almost embarrassingly *a prioristic* and unscientific for, of course, if one accepts standard empiricism, there can be no possible rationale for finding an empirically successful theory problematic, just because it does not fit in with one's "metaphysical predilections."

The attempt to pursue science in accordance with the edicts of standard empiricism actually plunges science, I have argued, into a measure of irrationality. This latent irrationality reveals itself most blatantly at moments of revolutionary change. Kuhn, in particular, has vividly described the manner in which rationality breaks down during phases of revolutionary change ([7], pp. 144–158). From the standpoint of aim oriented empiricism, we can see why it is that the widespread acceptance of standard empiricism is responsible for the breakdown in rationality, in the following way. A science has, let us suppose, some definite, but unarticulated blueprint B_1 . A succession of theories T_1, T_2, T_3 , are developed which successively move towards giving fuller and fuller articulations of B_1 . But then a new theory T_4 is developed, which is acceptable, and can be regarded as a definite improvement over T_3 , if a new blueprint B_2 is accepted, but which is unacceptable if B_1 is accepted. In order to judge rationally whether or not T_4 constitutes an advance over T_3 the crucial question that needs to be settled is whether the change of blueprint B_1 to B_2 constitutes an improvement or not. But it is precisely this crucial issue that scientists—unable to acknowledge and articulate scientific blueprints, due to their adherence to standard empiricism—are completely unable to discuss. The change of level 2 rules, of intelligibility criteria, brought about by the change of aims or blueprints B_1 to B_2 , cannot itself be rationally assessed. All that can happen is that one group of scientists which favors B_1 passionately declares T_4 to be unacceptable, while another group with equal passion, declares that T_4 constitutes a great advance. Rational discussion breaks down because orthodox ideas about science leave no room for the rational assessment of a change of blueprint and concomitant change of level 2 criteria of theory acceptability.

According to the aim oriented empiricist viewpoint advocated here, the very contrast between normal science and revolutionary science is an irrational bi-product of a failure to articulate and improve the aims of a science. For aim improvement, if pursued energetically, ought to anticipate, and in fact *provoke* revolutionary theoretical developments. In the above example, if aim improvement had been proceeding properly, B_2 would have been articulated and accepted long before T_4 was proposed; the arrival of T_4 would thus occasion no surprise or shock. Science pursued in accordance with aim oriented empiricism takes us beyond the Kuhnian dichotomy of normal and revolutionary science.

I have stressed the important implications that aim oriented empiricism has for

physics. Aim oriented empiricism has, however, perhaps even more important implications for the social sciences. For social scientists, anxious to make their disciplines scientifically respectable, have often sought (a) to suppress purely metaphysical “unscientific” speculation; (b) to model the *methods* of the social sciences as nearly as possible on the methods of the very successful natural sciences. According to aim oriented empiricism both these tendencies are disastrous. Without some kind of agreed aim or blueprint for a science, one hardly has a science at all. It is only when some kind of choice of blueprint has been made that one can have any idea of the kind of theory one is seeking to develop, and the kind of rules that ought to govern the acceptance and rejection of theories. The whole question of blueprint choice is especially important in a relatively young or unsuccessful science—as we may judge some of the social sciences to be. Furthermore, since sensible blueprints for social sciences are likely to be very different from any blueprint of physics, the level 2 methodology of the social sciences ought to be very different too. Thus, attempts to ape the methodology of the natural sciences are particularly unfortunate. All in all, anxious attempts to make the social sciences scientifically respectable by seeking to make them conform to the (false) ideals of standard empiricism will tend to have particularly unfortunate and stultifying consequences for these sciences.

REFERENCES*

- [1] Chiu, H. and Hoffmann, W. F. *Gravitation and Relativity*. New York: W. A. Benjamin, 1964.
- [2] Einstein, A. *Ideas and Opinions*. London: Alvin Redman, 1954.
- [3] Einstein, A. “Autobiographical Notes” and “Reply to Criticisms.” In *Albert Einstein: Philosopher-Scientist*. Edited by P. Schilpp. LaSalle, Illinois: Open Court, 1969.
- [4] Emmerson, J. *Symmetry Principles in Particle Physics*. Oxford: Clarendon Press, 1972.
- [5] Goodman, N. *Problems and Projects*. New York: Bobbs-Merrill Co., 1972.
- [6] Hoffmann, B. *Albert Einstein: Creator and Rebel*. London: Hart-Davis, MacGibbon, 1973.
- [7] Kuhn, T. S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1970.
- [8] Lakatos, I. “Falsification and the Methodology of Scientific Research Programmes.” In *Criticism and the Growth of Knowledge*. Edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press, 1970.
- [9] Maxwell, N. “Can there be Necessary Connections between Successive Events?” *British Journal for the Philosophy of Science* 19 (1968): 1–25.
- [10] Maxwell, N. “A New Look at the Quantum Mechanical Problem of Measurement.” *American Journal of Physics* 40 (1972).
- [11] Maxwell, N. “The Problem of Measurement—Real or Imaginary?” *American Journal of Physics* 41 (1973).
- [12] Maxwell, N. “Towards a Micro-Realistic Version of Quantum Mechanics.” *Foundations of Physics*. (forthcoming)
- [13] Penrose, R. and MacCallum, M. “Twistor Theory: An Approach to the Quantisation of Fields and Space-Time.” *Physics Reports* 6C (No. 4) (1973).
- [14] Popper, K. R. *The Logic of Scientific Discovery*. London: Hutchinson, 1959.
- [15] Riemann, G. “On the Hypotheses which Lie at the Foundations of Geometry.” In *A Source Book in Mathematics*. Edited by D. E. Smith. New York and London: McGraw-Hill, 1929.

* Reference citations in **Part II** are keyed to only this (**Part II**) list.

- [16] Swinburne, R. G. (ed.). *The Justification of Induction*. London: Oxford University Press, 1973.
- [17] Wheeler, J. A. *Geometrodynamics*. London: Academic Press, 1972.
- [18] Wheeler, J. A. "Superspace and the Nature of Quantum Geometrodynamics." In *Battelle Rencontre*. Edited by C. M. DeWitt and J. A. Wheeler. New York: W. A. Benjamin, 1968.
- [19] Wigner, E. P. *Symmetries and Reflections*. London: The M.I.T. Press, 1970.